REMARKS

I. Status of the Application

Claims 1, 2, 4-9 and 11-23 are presently pending in the application. Claims 19-23 stand withdrawn from consideration as being directed to non-elected subject matter. Claims 11-13, 15-19, 22 and 23 have been cancelled without prejudice to the filing of any appropriate continuation applications as being directed to non-elected subject matter. Claims 1, 2, 4-9 and 14 stand rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement. Claim 14 stands rejected under 35 U.S.C. § 112, first paragraph, as lacking enablement.

Applicants have amended the claims under consideration to more clearly define and distinctly characterize Applicants' novel invention. Claim 1 was amended to recite three Ets binding sites and one Sp-1 binding site, support for which can be found in the published application at least at Figure 1, where Applicants teach the presence of three Ets binding sites and one Sp-1 binding site in the -778 to -422 region of the EGP-2 promoter sequence; and to replace reference to Figure 1 with reference to the corresponding region of SEQ ID NO:5. Claim 14 was amended to recite a suicide gene, support for which can be found at least at paragraph [0019] of the published application, and to address formal matters. Claim 2 was amended to correct a typographical error and to address formal matters. Claim 20 was amended to depend from claim 5 and to address formal matters. Claim 21 was amended to recite "nucleic acid" and delete "method" in view of the amendment of claim 20 and to address formal matters. Claims 4-9 and 24 were amended to address formal matters.

Applicants respectfully submit that the amendments presented herein do not raise new issues requiring further search. Applicants respectfully request entry and consideration of the foregoing amendments, which are intended to place the case in condition for allowance, or at least in better condition for appeal.

II. Formal Matters

The Office Action, at page 2, asserts that claims 1, 2, 4-9 and 11-23 are pending in the application. However, Applicants respectfully submit that claims 1, 2, 4-9 and 11-24 are pending in the instant application. Claim 24 was added in Applicants' Supplemental Amendment mailed on May 9, 2006, a copy of which is provided herewith.

III. Restriction

The Office Action, at page 2, asserts that newly submitted claims 19-23 are directed to an invention that is independent or distinct from the invention originally claimed because they are drawn to a method of manufacture of a fusion construct, whereas the elected invention is an isolated nucleic acid. The Office Action states that the nucleic acid of Group I is patentably distinct from the claimed method of claims 19-23 because the nucleic acid of Group I can also be made by chemical synthesis, and accordingly, there would have been a search burden to examine these claims in a single application. Applicants respectfully traverse this restriction based on the amended claims now presented.

Without acquiescing to the instant restriction, Applicants respectfully submit that claims 19, 22 and 23 have been cancelled without prejudice to the filing of any appropriate continuation applications as being directed to non-elected subject matter. Claim 20 was amended to recite "a nucleic acid according to claim 5" and to delete reference to the method of cancelled claim 19. Claim 21 was amended to recite a "nucleic acid" and to delete "method." Thus, claims 20 and 21 ultimately depend from claim 1, and, therefore, pertain to elected subject matter.

Accordingly, Applicants respectfully request that the restriction of claims 20 and 21 be withdrawn.

IV. The Specification Provides Adequate Written Description for the Pending Claims

At page 3 of the instant Office Action, claims 1, 2, 4-9 and 14 stand rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement. Applicants respectfully traverse this rejection in view of the amended claims now presented. Applicants respectfully submit that the specification provides adequate written of the claimed nucleic acid sequence, and that one of skill in the art would readily understand the structure of the claimed nucleic acid sequence based on Applicants' teachings.

Without acquiescing to the rejection, Applicants respectfully submit that claim 1 has been amended to include the structural feature that at least three Ets binding sites and at least one Sp-1 binding site are present in the claimed promoter region. This claimed structural feature correlates with the claimed function of allowing expression of a nucleic acid of interest operably linked to the promoter in a cancer cell in an epithelium-selective manner. Applicants' novel invention is based in part on Applicants' elucidation of the basic minimal promoter region of the EGP-2 promoter that can bind the RNA pol II complex. Applicants have discovered that constructs containing the claimed nucleic acid sequence having the claimed binding sites are capable of mediating epithelial-specific expression (published application, paragraphs [0068] and [0076]).

The Office Action, at page 4, states that the claimed invention as a whole may not be adequately described where an invention is described solely in terms of a method of its making coupled with its function and there is no described or art-recognized correlation between the ***structure*** of the invention and its ***function***. The U.S. Patent and Trademark Office's Written

Description Guidelines state:

> The written description requirement for a claimed genus may be satisfied through sufficient description of a representative number of species by . . . disclosure of relevant, identifying characteristics, i.e., structure or other physical and/or chemical properties, by ***functional characteristics*** coupled with a ***known*** or ***disclosed correlation between function and structure,*** or by a combination of such identifying characteristics, sufficient to show the applicant was in possession of the claimed genus . . . .

66 Fed. Reg. 1099, 1106 (January 5, 2001), internal reference omitted, approved in *Enzo Biochem, Inc. v. Gen-Probe Incorporated*, 296 F.3d 1316, 1325, 63 U.S.P.Q.2d (BNA) 1609, 1613 (Fed. Cir. 2002) (emphasis added).

Applicants' claimed ***structural feature*** of having at least three Ets binding sites and at least one Sp-1 binding site in the claimed promoter region correlates with the claimed ***function*** of allowing ***expression*** of a nucleic acid of interest operably linked to the promoter in a cancer cell ***in an epithelium-selective manner***. These binding sites were ***well known*** in the art at the time of filing to have the function of regulating transcriptional activation in an epithelial cell-specific manner. In fact, Applicants' specification ***discloses*** that the Ets binding sites were known to modulate epithelium-specific transcription, and that that a combination of an Sp-1 binding site and an Ets binding site in close proximity was known in the art at the time of filing to regulate epithelium-specific gene expression (*See* published application, paragraph [0066]). Applicants have experimentally demonstrated that the claimed nucleic acid sequence could mediate expression of a nucleic acid of interest in an epithelial-specific manner (paragraph [0068]). Further, Applicants provide herewith Attachments A-D, review articles dated prior to Applicants' priority date, that describe Ets and Sp-1 transcription factors and their DNA binding regions (Berg (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:11109; Macleod et al. (1992) *Trends Biochem. Sci.* 17:251; Lania et al. (1997) *Int. J. Biochem. Cell Biol.* 29:1313; Suske (1999) *Gene* 238:291).

Applicants submit that at the time of filing, one of skill in the art of transcriptional regulation would readily understand that tissue-specific expression is not conferred by stretches of hundreds of nucleotides *per se*, but rather by the presence of combinations of short, *cis*-acting sequences capable of binding transcription factors. The skilled person would appreciate, based on the teachings of Applicants' specification, that not each and every individual nucleotide of the claimed nucleic acid sequence would be responsible for mediating transcriptional regulation. Instead, one of skill in the art would recognize that regulatory elements, such as the claimed Ets and Sp-1 binding sites identified within this region, would be critical for epithelial-specific expression.

Applicants' specification adequately describes the claimed nucleic acid sequence by providing SEQ ID NO:5 and describing several specific Ets binding site and Sp-1 binding site sequences (Figure 1). The specification provides protocols that may be used by one of skill in the art to determine whether a claimed nucleic acid sequence functions to mediate expression of a nucleic acid of interest in an epithelial-specific manner. The specification teaches how to construct the claimed nucleic acid sequence (paragraphs [0050]-[0053]), and how to express these constructs in a variety of cells (paragraphs [0042]-[0045]) and in transgenic mice (paragraphs [0054]-[0057] and [0070]). The specification describes an assay in which epithelial-specific expression may be detected (paragraph [0068]). Thus, one of skill in the art would understand that the specification sufficiently describes the claimed nucleic acid sequence.

The specification must be considered as a whole when determining whether the written description requirement is met. *In re Wright*, 866 F.2d 422, 425, 9 U.S.P.Q.2d (BNA) 1649, 1651 (Fed. Cir. 1989). The knowledge of one skilled in the art also must be considered, because the specification must "indicate[s] to persons skilled in the art that as of the [filing] date the applicant had invented what is now claimed." *All Dental Prodx LLC v. Advantage Dental*

Products Inc., 309 F.3d 774, 779, 64 U.S.P.Q.2d (BNA) 1945, 1948 (Fed. Cir. 2002). When read as a whole, taking into account the knowledge of persons skilled in the art at the filing date of the instant application, this specification indicates to those skilled in the art that Applicants had possession of the claimed subject matter at the time of filing. Accordingly, the Examiner is respectfully requested to reconsider and withdraw this rejection of claims 1, 2, 4-9 and 14 under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement.

V. Claim 14 is Enabled

At page 5 of the instant Office Action, claim 14 stands rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the enablement requirement. Applicants respectfully traverse this rejection based on the amended claim now presented.

35 U.S.C. § 112, first paragraph requires that the specification must enable a person skilled in the art to make and use the claimed invention. However, a specification need not, and should not, disclose what is well known in the art. The invention that one skilled in the art must be enabled to make and use is that defined by the claims of the particular application. The issue of adequate enablement depends on whether one skilled in the art could practice the claimed invention without undue experimentation. Enablement is not precluded by the necessity of some experimentation such as routine screening, even if it is extensive routine screening. Also, the fact that experimentation may be complex does not necessarily make it undue, if the art typically engages in such experimentation (MPEP 2164.01) if the level of skill in the art is high or if all of the methods needed to practice the claimed invention are well known. *In re Wands,* 8 U.S.P.Q. 2d 1400, 1406 (Fed. Cir. 1988).

> The determination of what constitutes undue experimentation in a given case requires the application of a standard of reasonableness, having due regard for the nature of the invention and the state of the art. (Citations omitted). The test is not

> merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed. *In re Wands*, 8 U.S.P.Q. 2d at 1404.

Without acquiescing to the rejection, Applicants respectfully submit that claim 14 was amended include the distinguishing feature that the medicament for ***treating epithelial cancer*** includes an isolated or recombinant nucleic acid sequence according to claim 1 having a nucleic acid sequence of interest that is a ***suicide gene***. Applicants respectfully submit that the instant specification provides ample direction and guidance to make and use the claimed medicament.

Determining whether a nucleic acid sequence has the claimed sequence identity, the claimed binding sites, and is expressed in an epithelium-specific manner would involve only ***routine screening***. For at least the reasons set forth above, Applicants respectfully submit that the instant specification teaches the nucleic acid sequence of SEQ ID NO:5, the sequences of Ets and Sp-1 binding sites, that these binding sites could modulate epithelial-specific gene expression, and that sequences of many such binding sites were known in the art at the time of filing. Applicants teach the use of suicide genes and their prodrugs, such as thymidine kinase/gancyclovir and cytosine deaminase/fluorocytosine, and that suicide genes were known in the art at the time of filing (paragraphs [0007] and [0019]). Accordingly, based on these teachings, one of skill in the art could easily make and use the claimed medicament.

Further, Applicants submit that using a combination of gene therapy and suicide genes to treat cancer was known in the art at the time of filing. Merely two months after Applicants' priority date of March 1, 2000, Springer and Niculescu-Duvaz published a review of the state of the art of suicide gene therapy (Attachment E, J. Clin. Invest. (May 2000) 105:1161, Applicants' priority date). Attachment E teaches that a gene from a foreign enzyme can be delivered and targeted to a tumor where it is to be expressed, and that an administered prodrug can be selectively activated by the foreign enzyme expressed in the tumor (page 1161, third paragraph). This review states that the use of transgenes to encode enzymes that activate specific prodrugs to create toxic products ("GDEPT") i.e., "suicide genes", "have already shown efficacy in vivo" (page 1161, second paragraph; page 1166, third full paragraph). In fact, gene therapy using suicide genes was being utilized in ***more than 25 clinical trials*** at the time this review was published (Table 1, page 1162).

The Office Action, at page 5, states that immune response is an obstacle of gene therapy and indicates that the claims are not limiting the medicament with a DNA-based vector or therapeutic genes that would not elicit an immune response. The Springer and Niculescu-Duvaz review, however, does not report host immune system rejection of suicide transgenes. Instead, this reference indicates that the immune response ***increases the efficacy*** of gene therapy (paragraph bridging pages 1165-66 and first full paragraph of page 1166).

For at least these reasons, Applicant's specification, coupled with the level of skill in the art, enables a person of skill in the art to make and/or use the claimed invention. Accordingly, the Examiner is respectfully requested to reconsider and withdraw the rejection of claim 14 under 35 U.S.C. § 112, first paragraph, as lacking enablement.

VI. CONCLUSION

Having addressed all outstanding issues, Applicants respectfully request reconsideration and allowance of the case. To the extent the Examiner believes that it would facilitate allowance of the case, the Examiner is requested to telephone Applicants' attorney at the number below.

Respectfully submitted,

Dated: September 5, 2006

J Stone-Hulslander
Judith L. Stone-Hulslander, Reg. No. 55,652
BANNER & WITCOFF, LTD.
28 State Street, 28th Floor
Boston, MA 02109
(617) 720-9600

A

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 11109–11110, December 1992

Commentary

Sp1 and the subfamily of zinc finger proteins with guanine-rich binding sites

Jeremy M. Berg

Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

The Cys$_2$His$_2$ zinc finger proteins are a class of DNA binding proteins that contain sequences of the form (Tyr,Phe)-Xaa-Cys-Xaa$_{2,4}$-Cys-Xaa$_3$-Phe-Xaa$_5$-Leu-Xaa$_2$-His-Xaa$_{3-5}$-His, usually in tandem arrays (1–4). Each of these sequences binds a zinc(II) ion to form a structural domain termed a zinc finger. Structural studies by NMR (5–11) and x-ray crystallography (12) have revealed that these domains adopt very similar structures that consist of a β hairpin followed by a helix. Yet, because of variations of certain key amino acids from one zinc finger to the next, each domain makes its own unique contribution to DNA binding affinity and specificity.

The most well-understood members of this class of proteins are a subset that bind to relatively guanine-rich binding sites. This subset includes Spl (13), the Zif268/NGFI-A/Krox-20,24/Egr1,2/Wilm tumor family (14–19) and yeast ADR1 (20). Kriwacki *et al.* (21) report studies of the DNA binding domain of Spl that consists of three zinc finger domains. They demonstrated three major points. (*i*) They showed that a 92-amino acid peptide corresponding to the zinc finger region of Spl with very little flanking sequence specifically bound DNA containing sites that approximated the Spl binding site 5′-GGGGCGGGGC-3′. (*ii*) They demonstrated the ability of this peptide to distinguish between different DNA sequences in a manner reminiscent of that of intact Spl. (*iii*) They showed that the 5′-GGGGCG region of the binding site contributed more to the overall binding affinity than does the GGGC-3′ region.

These observations can be interpreted in light of other reports. Several groups have demonstrated that fragments of Spl that included the zinc finger domains showed DNA binding properties very similar to those of the intact protein (22–26). The smaller fragment studied by Kriwacki *et al.* (21) is the smallest polypeptide studied to date beginning only 7 residues before the first Cys residue and extending only 3 amino acids past the last His residue.

The crystallographic studies of the three zinc finger domains of Zif268 bound to a cognate oligonucleotide have provided the most detailed structural information to date concerning zinc finger protein–DNA recognition (12). Three amino acids per domain play the most direct role in determining site preference. These are indicated as $\mathbf{Xaa}^{A}$, $\mathbf{Xaa}^{B}$, and $\mathbf{Xaa}^{C}$ in the sequence below:

$$\text{(Tyr,Phe)-Xaa-Cys-Xaa}_{2,4}\text{-Cys-Xaa}_{3}\text{-}$$
$$\text{Phe-Xaa-}\mathbf{Xaa}^{A}\text{-Xaa-Xaa-}\mathbf{Xaa}^{B}\text{-}$$
$$\text{Leu-Xaa-}\mathbf{Xaa}^{C}\text{-His-Xaa}_{3-5}\text{-His.}$$

The first and third fingers of Zif268 have $\mathbf{Xaa}^{A}$ = Arg, $\mathbf{Xaa}^{B}$ = Glu, $\mathbf{Xaa}^{C}$ = Arg. Each of these binds the triplet GCG with the two Arg residues forming a pair of hydrogen bonds with guanine and the Glu interacting indirectly with cytosine. Note that the Arg in $\mathbf{Xaa}^{A}$ interacts with the guanine at the 3′ end of the triplet. The second domain of Zif268 has $\mathbf{Xaa}^{A}$ = Arg, $\mathbf{Xaa}^{B}$ = His, $\mathbf{Xaa}^{C}$ = Thr. This finger interacts with the sequence 5′-TGG-3′ in the cocrystal structure via two hydrogen bonds from the Arg to the 3′ guanine and one hydrogen bond between the His and N7 of the central guanine.

The three zinc finger domains (shown, in order, fingers 1–3) of Spl have potential contact residues

`Xaa^A = Lys,Xaa^B = His,Xaa^C = Ala`

`Xaa^A = Arg,Xaa^B = Glu,Xaa^C = Arg`

`Xaa^A = Arg,Xaa^B = His,Xaa^C = Lys.`

Assuming that the corresponding contacts from the Zif268 structure apply to Spl, the binding site can be predicted to be 5′-NGG-GCG-NGN-3′, where the 5′-NGG triplet is contacted by the third finger, the central GCG is contacted by the second finger, and the NGN triplet is contacted by the first finger. The His residues in position $\mathbf{Xaa}^{B}$ are likely to bind well to adenine as well as guanine since the observed hydrogen bond involved N7. This has been directly demonstrated for an Spl variant with $\mathbf{Xaa}^{B}$ = His in the second domain (27). Thus, the predicted site can be written 5′-N(G/A)G-GCG-N(G/A)N-3′. This clearly is a good approximation to the Spl consensus (23, 28). Moreover, it accounts for the many aspects of the variations in the known Spl binding sites as well as the decreased affinity of the Spl peptide for two mutant sites reported by Kriwacki *et al.* (21). Finally, this analysis provides a rationalization for the asymmetry of the contributions to overall binding affinity noted by Kriwacki *et al.* since the first domain (which contacts the 3′ end of the binding site) has only $\mathbf{Xaa}^{B}$ = His predicted to make a directly precedented contact. While it is likely that $\mathbf{Xaa}^{A}$ = Lys also hydrogen bonds to a base, $\mathbf{Xaa}^{C}$ = Ala is obviously incapable of such an interaction.

This observation raises one of the remaining puzzles about Spl–DNA interactions. If $\mathbf{Xaa}^{C}$ = Ala does not interact directly with the DNA, why is the underlined guanine in 5′-GGG-GCG-<u>G</u>GG so well conserved in the known Spl binding sites? This position is a thymine in several of the known high-affinity sites but no direct studies have been reported that bear on the effects of mutations at this position on Spl binding affinity. It is possible that changes in this site do not affect affinity significantly or that guanine is preferred but is recognized by a more indirect mechanism.

These studies illustrate one approach to investigating specific protein–DNA interactions—namely, examination of the ability of a given protein to distinguish between different binding sites. An alternative approach is to investigate the effects of changes in the protein sequence on DNA binding affinity and specificity. Change-of-specificity mutants (in which DNA binding ability is not simply disrupted but changed to a different preferred DNA sequence) are particularly useful. The first such mutants were reported by Nardelli *et al.* (29) for Krox-20, a close homologue of Zif268, with three zinc finger domains and the same contact residues. The wild-type protein was shown to bind 5′-GCG-GGG-GCG-3′ but not 5′-GCG-GCG-GCG-3′. In the second domain, $\mathbf{Xaa}^{B}$ was changed from His to Glu and $\mathbf{Xaa}^{C}$ was changed from Thr to Arg. The double mutant showed the reverse specificity for the two DNA sequences above. Obviously, the two mutations put the same contact residues in the second domain that were already found in the first and third domains. Thus, binding to 5′-GCG-GCG-GCG-3′ is not surprising but it is important to realize that these experiments were published prior to the report of the Zif268 cocrystal structure. Other mutants studied resulted in reduction in the selectivity of DNA binding.

A more extensive series of change-of-specificity mutants of Sp1 itself has been developed by Desjarlais and Berg (25–27). These involved changes in the second zinc finger domain. The first of these changed the binding specificity from 5′-GGG-GCG-GGG-3′ to 5′-GGG-GCT-GGG-3′ via mutation of **Xaa**$^{\mathbf{A}}$ from Arg to Gln and **Xaa**$^{\mathbf{B}}$ from Glu to Asp as well as one additional change in a residue that interacts with the **Xaa**$^{\mathbf{A}}$ side chain (25, 26). The fact that three amino acid sequence changes were required to change the preferred binding site in a single position illustrates the necessity to consider the entire recognition helix; only certain sets of contact residues are mutually compatible. An additional mutant was produced by changing **Xaa**$^{\mathbf{B}}$ from Asp to Asn in the above context. This mutant binds to 5′-GGG-GAT-GGG-3′. A single change of **Xaa**$^{\mathbf{B}}$ from Glu to His in the wild-type Sp1 background bound to 5′-GGG-G(A/G)G-GGG-3′. This is the inverse of one of the Krox-20 mutants discussed above. Finally, a mutant with **Xaa**$^{\mathbf{B}}$ changed from Glu to Leu bound 5′-GGG-G(A/C/T)G-GGG-3′.

A final member of this protein subfamily that has been extensively studied is yeast ADR1. This protein has two zinc fingers in its DNA binding domain and it generally binds as a dimer on sites with approximate dyad symmetry (30). The two putative contact residues (shown, in order, fingers 1 and 2) are

$\text{Xaa}^{A} = \text{Arg}, \text{Xaa}^{B} = \text{His}, \text{Xaa}^{C} = \text{Arg}$

$\text{Xaa}^{A} = \text{Arg}, \text{Xaa}^{B} = \text{Leu}, \text{Xaa}^{C} = \text{Arg}.$

One of the preferred binding sites is 5′-TTG-GAG-3′. These contacts are consistent with those above except that the Arg in position **Xaa**$^{\mathbf{C}}$ in the second finger abuts thymine rather than guanine. Three change-of-specificity mutants were produced via an extensive study of mutations in the first zinc finger (31). Changing **Xaa**$^{\mathbf{A}}$ from Arg to Gln produced a protein that preferred the site 5′-TTG-GAA-3′. It is interesting to note that the preference associated with the **Xaa**$^{\mathbf{A}}$ = Gln differs from that observed in the Sp1 mutant above. Glutamine is notable for its ability to interact with DNA bases in different ways; in the cocrystal structure of 434 repressor bound to DNA, three Gln residues make three different types of contact with DNA (32). A second mutant involved changing **Xaa**$^{\mathbf{B}}$ from His to Thr. This mutant protein preferred 5′-TTG-GCG-3′. Finally, a mutant involving a change of **Xaa**$^{\mathbf{C}}$ from Arg to Asn preferred 5′-TTG-AAG-3′. These results on ADR1 confirm and extend the results from the Sp1 mutants.

Because of the apparent modular nature of the zinc finger domains in proteins of this class, it has been very tempting to hope that codes could be deduced that interrelate the amino acid sequences of zinc finger domains and the sequences of their preferred binding sites. This would allow prediction of binding site sequences for naturally occurring zinc finger proteins and design of zinc finger proteins with preselected DNA binding properties. As the above discussion indicates, considerable progress toward this goal has been made for this small subset of zinc finger proteins that interact with guanine-rich binding sites. However, it is important to note that many zinc finger proteins have amino acids in the potential contact positions whose interactions with DNA have not been elucidated. Furthermore, some zinc finger proteins such as *Drosophila* hunchback have been shown to interact with very (A+T)-rich binding sites (33, 34). Future studies should reveal whether similar rules can be developed for these protein–DNA interactions.

1. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609–1614.
2. Klug, A. & Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464–470.
3. Evans, R. M. & Hollenberg, S. M. (1988) *Cell* **52**, 1–3.
4. Berg, J. M. (1990) *Annu. Rev. Biophys. Biophys. Chem.* **19**, 405–421.
5. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245**, 635–637.
6. Párraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T. & Klevit, R. E. (1988) *Science* **241**, 1489–1492.
7. Neuhaus, D., Nakaseko, Y., Nagai, K. & Klug, A. (1990) *FEBS Lett.* **262**, 179–184.
8. Kochoyan, M., Keutmann, H. T. & Weiss, M. A. (1991) *Biochemistry* **30**, 9396–9402.
9. Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. & Gronenborn, A. M. (1990) *Biochemistry* **29**, 9324–9334.
10. Omichinski, J. G., Clore, G. M., Robien, M., Sakaguchi, K., Appella, E. & Gronenborn, A. M. (1992) *Biochemistry* **31**, 3907–3917.
11. Krizek, B. A., Amann, B. T., Kilfoil, V. J., Merkle, D. L. & Berg, J. M. (1991) *J. Am. Chem. Soc.* **113**, 4518–4523.
12. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809–817.
13. Kadonaga, J. T., Carner, C., Masiarz, F. R. & Tjian, R. T. (1987) *Cell* **51**, 1079–1090.
14. Christy, B. A., Lau, L. F. & Nathans, D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7857–7861.
15. Milbrandt, J. (1987) *Science* **238**, 797–799.
16. Chavrier, P., Vesque, C., Galliot, B., Vigneron, M., Dollé, P., Duboule, D. & Charnay, P. (1990) *EMBO J.* **9**, 1209–1218.
17. Joseph, L. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7164–7168.
18. Gessler, M., Poustka, A., Cavenee, W., Neve, R. L., Orkin, S. H. & Bruns, G. A. P. (1990) *Nature (London)* **343**, 774–778.
19. Haber, D. A., Buckler, A. J., Glaser, T., Call, K. M., Pelletier, J., Sohn, R. L., Douglass, E. C. & Housman, D. E. (1990) *Cell* **61**, 1257–1269.
20. Hartshorne, T. A., Blumberg, H. & Young, E. T. (1986) *Nature (London)* **320**, 281–287.
21. Kriwacki, R. W., Schultz, S. C., Steitz, T. A. & Caradonna, J. P. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9759–9763.
22. Courey, A. J. & Tjian, R. (1988) *Cell* **55**, 887–898.
23. Thiesen, H.-J. & Bach, C. (1990) *Nucleic Acids Res.* **18**, 3203–3209.
24. Kuwahara, J. & Coleman, J. E. (1990) *Biochemistry* **29**, 8627–8631.
25. Desjarlais, J. R. & Berg, J. M. (1992) *Proteins: Struct. Funct. Genet.* **12**, 101–104.
26. Desjarlais, J. R. & Berg, J. M. (1992) *Proteins: Struct. Funct. Genet.* **13**, 272.
27. Desjarlais, J. R. & Berg, J. M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7345–7349.
28. Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23.
29. Nardelli, J., Gibson, T. J., Vasque, C. & Charnay, P. (1991) *Nature (London)* **349**, 175–178.
30. Eisen, A., Taylor, W. E., Blumberg, H. & Young, E. T. (1988) *Mol. Cell. Biol.* **8**, 4552–4556.
31. Thukrel, S. K., Morrison, M. L. & Young, E. T. (1992) *Mol. Cell. Biol.* **12**, 2784–2792.
32. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) *Science* **242**, 899–907.
33. Stanojevic, D., Hoey, T. & Levine, M. (1989) *Nature (London)* **341**, 331–335.
34. Treisman, J. & Desplan, C. (1989) *Nature (London)* **341**, 335–337.

B

The *ets* gene family

Kay Macleod, Dominique Leprince and Dominique Stehelin

Ets proteins have a conserved DNA-binding domain and regulate transcriptional initiation from a variety of cellular and viral gene promoter and enhancer elements. Some members of the Ets family, Ets-1 and Ets-2, cooperate in transcription with the AP-1 transcription factor, the product of the proto-oncogene families, *fos* and *jun*, while others, Elk-1 and SAP-1, form ternary complexes with the serum response factor (SRF). Certain *ets* gene family members possess transforming activity while others are activated by proviral integration in erythroleukaemias.

THE v-*ets* ONCOGENE, the first member of the *ets* gene family to be described, was discovered in the E26 ('E-*t*wenty-*s*ix') acutely transforming retrovirus of chicken, from which it derives its name[1,2]. Other family members cloned from a variety of organisms[1–19], are listed in Table I. In most family members, the conserved DNA-binding domain (the ETS domain), is localized at the carboxyl terminus of the protein, with the exception of Elk-1, SAP-1 and Elf-1, where it is found at the amino terminus[9,10,16,17]. The ETS domain, which covers approximately 85 amino acids (Fig. 1), has no structural homology to other known DNA-binding motifs such as the zinc finger, homeodomain, leucine zipper or helix-turn-helix motifs[6] but has limited homology to the DNA-binding domain of the c-*myb* oncogene product. In c-Myb, nine Trp residues are repeated every 18 or 19 amino acids around a predicted α-helical structure[6], whereas the ETS domain contains only two or three conserved Trp residues (Fig. 1)[6].

K. Macleod, **D. Leprince** and **D. Stehelin** are at CNRS, Unite d'Oncologie Moleculaire, Institut Pasteur, 1 rue Calmette – BP 245, 59019 Lille Cédex, France.

The sequence conservation of the proteins and the chromosome locations of the genes provide clues to gene duplication and divergence events that gave rise to this gene family. For example, *ets-1* and *fli-1* both map to the same region of human chromosome 11q23, while *ets-2* and *erg* both map to 21q22 (Refs 2,11). However, the homology between *ets-1* and *ets-2* and between *erg* and *fli-1* is greater than the homology between those genes that map next to each other on the same chromosome[2,11]. This suggests that duplication of an ancestral *ets*-related gene may have occurred, giving rise to two sister genes, which subsequently diverged over time. This was followed by duplication of the entire locus to another chromosome.

Regulation of *ets* expression and activity

The expression of *ets* genes is controlled at several levels: the initiation of

```
                 ** *** **         * *       *B      *     **  B  B  ***    B * ***      B       * *
Ets-1     GSGP-IQLWQFLLELLTDKS-CQSFISWT-GDGWEFKLSDPDEVA--RRWGKRK-NKPKMNYEKLSRGLRYYYDKNIIHKTAGKRYVYRFV
Ets-2     GSGP-IQLWQFLLELLSDKS-CQSFISWT-GDGWEFKLADPDEVA--RRWGKRK-NKPKMNYEKLSRGLRYYYDKNIIHKTSGKRYVYRFV
GABP-α    NNGQ-IQLWQFLLELLTDKD-ARDCISWVGDEEGEFKLWQPELVA--QKWGQRK-NKPTMNYEKLSRALRYYYDGDMICKVQGKRFVYKFV
Erg       GSGQ-IQLWQFLLELLSDSS-NSSCITWE-GTNGEFKMTDPDEVA--RRWGERK-SKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD
Fli-1     GSGQ-IQLWQFLLELLSDSA-NASCITWE-GTNGEFKMTDPDEVA--RRWGERK-SKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD
D-Elg     NNGQ-VQLWQFLLEILTDCE-HTDVIEWV-GTEGEFKLTDPDRVA--RLWGEKK-NKPAMNYEKLSRALRYYYDGDMISKVSGKRFAYKFD
E74A      GST--TYLWEFLLKLLQDREYCPRFIKWTNREKGVFKLVDSKAVS--RLWGMHK-NKPDMNYETMGRALRYYYQRGILAKVDGQRLVYQFV
Elf-1     GNT--IYLWEFLLALLQDKATCPKYIKWTQREKGIFKLVDSKAVS--RLWGKHK-NKPDMNYETMGRALRYYYQRGILAKVEGGRLVYQF-
SAP-1     -MDSAITLWQFLLQLLQKPQ-NKHMICWTS-MDGQFKLLQAEEVA--RLWGIRK-NKPNMNYDKLSRALRYYYVKNIIKKVNGQKFVYKFV
Elk-1     PS---VTLWQFLLQLLREQG-NGHIISWTSADGGEFKLVDAEEVA--ALWGLAK-NKTNMNYDKLSRALAYYYDKNIIIAKVSGQKFVYKV
PEA3      QRRGALQLWQFLVALLDDPT-NAHFIAWTG-RGMEFKLIEPEEVA--RLWGIQK-NRPAMNYDKLSRSLRYYYEKGIMQKVAGERYVYKFV
PU-1      GSKKKIRLYQFLLDLLRSGD-MKDSIWWVDKDKGTFQFSSKHKEALAHRWGIQKGNRKKMTYQKMARALRNYG-KTGEVKKVKKKLTYQFS
```

Figure 1

Alignment of the ETS domain primary amino acid sequences. The conserved Trp residues (W) are boxed. Other highly conserved amino acids are indicated by asterisks (*), while conserved basic residues are indicated by (B).

© 1992, Elsevier Science Publishers, (UK) 0376–5067/92/$05.00

transcription[22], alternative splicing[4,5], post-translational modification[7] and protein stability[23]. The expression patterns of different members of the *ets* gene family vary between tissues (Table I).

In T cells, expression of *ets-1* is highest in quiescent cells and decreases following T-cell activation[20]. Conversely, *ets-2* gene expression increases from a low level following T-cell activation and reaches maximum levels 1–2 h after induction[20]. The mechanism of coordination of the expression of these two genes during T-cell differentiation is unclear.

Regulation of transcriptional initiation is best understood for the human *ets-2* gene. Like all other *ets* promoters described, and many other oncogene promoters, the *ets-2* promoter lacks the classical TATA and CCAAT box sequences found in many RNA polymerase II-transcribed genes. A putative Ets-binding site (EBS) has been identified downstream of the promoter (Table II) and *ets-2* may be negatively autoregulated. Adjacent to this EBS, there is a putative activator protein 1 (AP-1) site and factors bound at these sites cooperate in *trans*-activation. These EBS and AP-1 sites bind activator and repressor factors in a cell-specific fashion[22].

The promoter of the human *ets-1* gene contains putative AP-1, EBS, AP-2 and Sp1 sites. Expression of exogenous AP-2 and Ets-1 increases the level of transcription from the *ets-1* promoter, implying an activating role for AP-2 in *ets-1* expression and positive autoregulation of *ets-1* transcription[3,24].

Chicken and human *ets-1* RNAs are alternatively spliced[4,5]. In chicken, two forms of Ets-1 mRNA are generated by alternative splicing of 5' exons: α and β

Table I. The *ets* gene family

Protein	Source	Molecular mass (kDa)	Amino acid homology to ETS domain of Ets-1 (%)	Human chromosomal location	Expression/features
Ets-1	human mouse chicken	39–52 63 54/68	100	11q23	Elevated expression in thymus and endothelial cells; phosphorylated; alternatively spliced; positively autoregulates transcription[1–6].
Ets-2	human chicken	58/62	90	21q22	Expression induced following macrophage differentiation and T-cell activation; alternatively spliced; phosphorylated[2,6,7].
Erg	human	41/52	70	21q22	Alternatively spliced; 98% homologous to Fli-1[6,8].
Fli-1	human mouse	 51	68	11q23	Activated by proviral insertion of Friend MuLV; 98% homologous to Erg[11].
Elk-1	human	60	76	Xp11.2	ETS domain located in the amino terminus of the protein; forms ternary complex with SRF; shows three regions of homology with SAP-1[6,9,10].
SAP-1 a/b	human	58/52	75	ND	SRF accessory protein 1, which, like Elk-1, forms a ternary complex with SRF over the *c-fos* SRE; contains three regions of homology to Elk-1, including the ETS domain, which is located in the amino terminus of the protein; the two isoforms, SAP-1a and SAP-1b, differ in their carboxyl termini[17].
Spi-1/ PU1	human mouse	30	38	11p11.22	Activated in Friend erythroleukaemia by proviral insertion of SFFV; normal expression of the PU-1 transcription factor is restricted to B cells and macrophages[6,12].
E74A/B	*Drosophila*	110/120	50	*Drosophila* chromosome 3L74EF	E74A is induced by ecdysone and regulates the expression of E74B, which is also *ets*-related[6].
Elf-1	human	68	50	ND	The ETS domain is the human homologue of the E74A protein of *Drosophila*; binds to the NF-AT and NFIL-2B sites in the interleukin-2 promoter and the human immunodeficiency virus 2 LTR[16].
GABP-α	rat	51	82	ND	High-level expression in rat thymus; complexes with GABP-β, which contains ankyrin repeats, and is related to the Notch protein[13,14].
D-Elg	*Drosophila*	15	64	*Drosophila* chromosome 3R97D	Contains only a DNA-binding domain; maternally expressed message and also expressed throughout embryogenesis[15].
PEA3	mouse	68	63	ND	Expressed in mouse brain and epididymis and in fibroblast and epithelial cell lines; down-regulated in embryonic cell lines in response to retinoic-acid-induced differentiation[18].
TCF1-α	human	55	ND	ND	Very limited homology to ETS domain exists within the HMG box of this factor; expression is restricted to the thymus and is induced following T-cell activation; regulates activity of the TCRα enhancer[19].

Abbreviations: ND, not determined; MuLV, murine leukaemia virus; SFFV, spleen focus forming virus; SRF, serum response factor; NF-AT, nuclear factor of activated T cells; NFIL-2B, nuclear factor of interleukin 2B; LTR, long terminal repeat; GABP-β, GA-binding protein; PEA3, polyomavirus enhancer activator 3; HGM, high mobility group; TCR, T-cell receptor.

Table II. Regulatory elements containing Ets-binding sites

Regulatory element	EBS sequence	Ets protein known to bind
ets-2 promoter[22]	TGGA**GGAA**GT	Ets-1/Ets-2
Interleukin 2 enhancer (NF-AT-1)[13]	AAGGA**GGAAA**	Elf-1
polyomavirus enhancer (PEA3)[27]	AGCA**GGAA**GT	Ets-1/Ets-2/Erg/Elk-1
Moloney Sarcoma Virus LTR[32]	GAGC**GGAA**GC	Ets-1/Ets-2
SV40 enhancer[12]	AAGA**GGAA**CT	PU-1
ICP4 promoter[13,14]	6X (C**GGAA**A/G)[a]	GABP-α (binding depends on GABP-β)
HTLV I LTR[31]	GGA**GGAA**AT	Ets-1/Ets-2
HTLV I LTR[31]	CCG**GGAA**GC	Ets-1/Ets-2
stromelysin 1 promoter[28]	GCA**GGAA**GC	Ets-1/Ets-2
stromelysin 1 promoter[28]	CCA**GGAA**AT	Ets-1/Ets-2
c-*fos* promoter[10]	CA**GGAT**GT	SAP-1 (binding depends on SRF)
T-cell receptor α enhancer[33]	CAGA**GGAT**GT	Ets-1

Abbreviations: EBS, Ets-binding site; HTLV, human T-cell leukaemia virus.
[a] X = any nucleotide.

of p68 are replaced by I^{54} in p54 (Fig. 2). Unlike p54, p68 is expressed in endothelial cells rather than in the thymus or T cells[4]. The differential functions of these proteins are thought to be associated with the different *trans*-activation domains encoded by the 5' exons[25]. Exons α and β have not been identified in any mammalian species (O. Albagli, pers. commun.), but in human *ets-1*, alternative splicing of exons IV and VII has been observed[5]. Exon VII encodes a domain that has a negative regulatory effect on the DNA-binding domain, which may be the result of phosphorylation of Ser residues in this region[5]. Ets-2 also has this inhibitory domain but PU-1 lacks such a domain[46].

Phosphorylation of both chicken and human Ets-1 and Ets-2 results in the loss of non-specific DNA-binding activity[23]. The cellular kinases responsible for phosphorylating Ets-1 and Ets-2 proteins are unknown, although putative sites of phosphorylation for protein kinase C, casein kinase II and calmodulin kinase have been identified. Following activation of T cells, in addition to the reduced level of *ets-1* transcription, there is an increase in phosphorylation of Ets-1 that results in reduced DNA-binding activity[23]. The half-life of Ets-2 increases following T-cell activation from approximately 20 min to greater than 2 h as a consequence of post-translational modification[26]. Protein kinase C (PKC) inhibitors block this effect, suggesting that PKC directly or indirectly modifies the stability of Ets-2 (Ref. 26).

PEST sequences (regions rich in Pro, Glu, Ser and Thr), may play a role in protein turnover and are found in both the transcription factor PU-1 (from amino acids 121–159)[12] and $p68^{c\text{-}ets\text{-}1}$ (from amino acids 295–326). The PEST sequences are located between putative *trans*-activation domains and DNA binding domains and cleavage may create a DNA-binding domain free from the regulatory restraints imposed by amino-terminal domains.

The Ets-binding site

The $p68^{c\text{-}ets\text{-}1}$ protein binds to the PEA3 element of the polyoma virus enhancer[27]. The DNA sequence AGCAGGAAGT is specifically recognized by $p68^{c\text{-}ets\text{-}1}$ in the polyoma enhancer, and binds $p68^{c\text{-}ets\text{-}1}$ with moderate affinity as determined by electrophoretic mobility shift assay (EMSA). Mutational analysis of the PEA3 element has identified several higher-affinity EBS, such as AGCCGGAAGT[28,29], in which the second A has been replaced by C. Selection and amplified binding studies identified further mutations that increase the binding affinity of the EBS for $p54^{c\text{-}ets\text{-}1}$, such as a G→A transition in addition to the A→C transversion to generate AACCGGAAGT[30]. $p54^{c\text{-}ets\text{-}1}$, $p58^{c\text{-}ets\text{-}2}$, Erg and Elk-1 show similar affinities for the same sequences as $p68^{c\text{-}ets\text{-}1}$ (Refs 8,9,28) while $p30^{spi\text{-}1}$ and Elf-1 appear to have different sequence specificities[16,28].

EBS have been identified in other cellular and viral gene regulatory sequences (Table II) including the long terminal repeats (LTR) of the human T-cell leukaemia virus I (HTLV-I)[31], the Moloney Sarcoma Virus (MSV) LTR[32] and the rat stromelysin 1 promoter[28]. A combination of EMSA and methylation interference studies on the binding of Ets proteins to the MSV[32] and HTLV-I[31] LTRs has shown that the GGAA purine core is essential for the specific binding of Ets-related proteins to these sequences. However, this is not the case for sites within certain T-cell-specific enhancers, such as the enhancers of the genes encoding T-cell receptor-α(TCR-α)[33] and -β(TCR-β)[17]. In these cases, an A→T transversion in the purine core to generate GGAT is frequently observed (Table II).

When viral point mutations and the carboxyl terminus of the v-*ets* protein were introduced into a 35 kDa version of the $p68^{c\text{-}ets\text{-}1}$ protein, the altered protein lost its ability to bind to the EBS[34]. This suggested that the v-*ets* protein has either a different DNA-binding specificity to $p68^{c\text{-}ets\text{-}1}$ or does not bind DNA at all[34]. Conflicting evidence has emerged showing that the v-*ets* protein has a stronger binding affinity for the EBS than the full-length $p68^{c\text{-}ets\text{-}1}$ protein and that cleavage of c-Ets-1 generates a truncated protein that binds strongly to the EBS[35]. It has been suggested that there is a negative regulatory domain in the amino terminus of Ets-1 that interacts with the carboxyl terminus of the protein to prevent DNA binding. Cleavage of the protein would repress DNA-binding by $p68^{c\text{-}ets\text{-}1}$ (Ref. 35). The v-Ets protein has a different carboxyl terminus, so it has been proposed that the negative regulatory domain present in this viral protein would be redundant, since it would be unable to associate with carboxy-terminal sequences which are present only in the cellular protein[35]. It will be interesting to determine whether the viral fusion protein active in erythroleukaemia, $p135^{\delta gag\text{-}myb\text{-}ets}$, binds to the EBS or not.

Ets target genes

The identification of EBS in different promoters/enhancers has permitted identification of genes that may be regulated by Ets-related proteins; the ex-

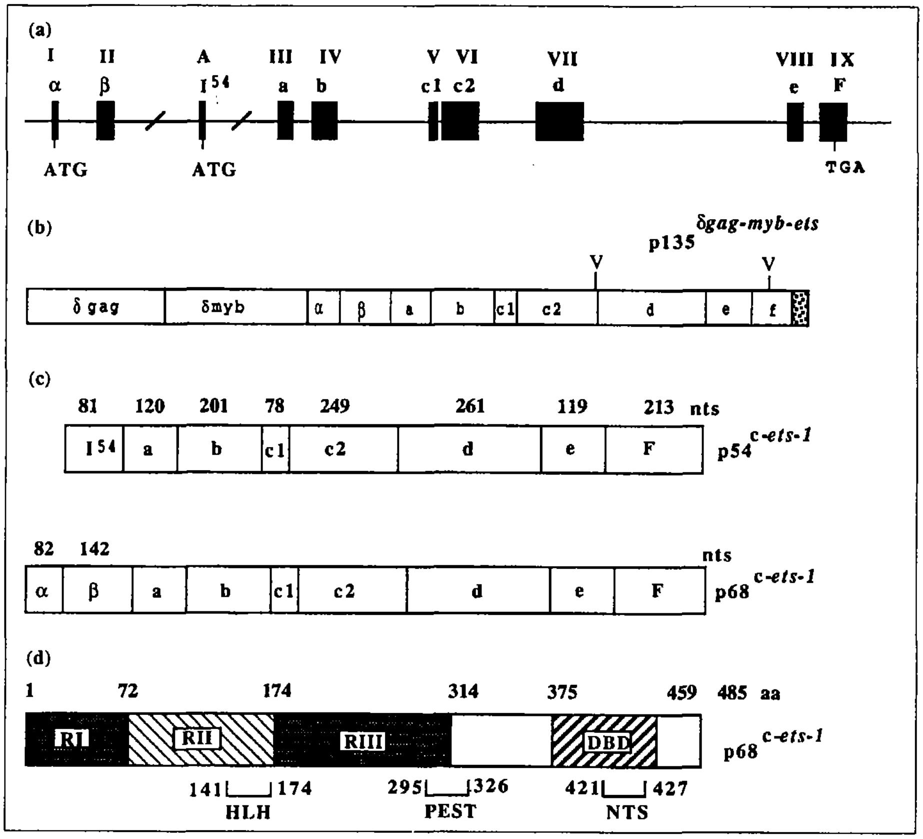


Figure 2

(a) Genomic organization of the chicken *c-ets-1* locus. Chicken exons (α, β, I^{54}, a, b, c1, c2, d, e and F) are shown as black boxes, along with an alternative nomenclature used for the chicken and human exons (I, II, A, III, IV, V, VI, VII, VIII, IX). Two translation start signals (ATG) are found in exons α and I^{54}, while the stop signal (TGA) is in exon F. **(b)** The 135 kDa viral fusion protein of the E26 virus showing the organization of *ets* sequences relative to those of *gag* and *myb*. The α and β exons, but not the I^{54} exon, have been transduced. The positions of non-conservative mutations (V) in *v-ets*, compared with *c-ets-1* [see (c)] and a change in carboxy-terminal sequences (▨) are shown. **(c)** The organization and sizes of the different exons in the alternatively spliced *c-ets-1* mRNAs that give rise to $p54^{c\text{-}ets\text{-}1}$ and $p68^{c\text{-}ets\text{-}1}$. The number of nucleotides (nts) in each exon is indicated. **(d)** Functional domains of $p68^{c\text{-}ets\text{-}1}$. The amino acid residues are numbered and aligned with the nucleotide sequences of the p68 mRNA in (c) above, thus indicating the exons that code for each functional domain of the protein. RI, region I (*trans*-activation domain); RII, region II (regulatory domain); RIII, region III (*trans*-activation domain); DBD, DNA-binding domain; HLH, helix-loop-helix domain; PEST, region rich in Pro, Glu, Ser and Thr residues which may target protein cleavage to this region of the protein; NTS, nuclear targeting signal.

pression of these genes often correlates with the expression of Ets-related proteins. A significant number of lymphoid-specific genes contain EBS in their regulatory sequences, including the enhancers of the T-cell receptor-α[33] and the T-cell receptor-β (M. Owen, pers. commun.). The EBS of the TCR-β enhancer mediates both basal and phorbol ester/T-cell activator-induced expression of the gene encoding TCR-β. Expression of Ets-1 represses both basal and inducible activity of the TCR-β enhancer, while expression of Ets-2 has no effect on enhancer activity. Derepression of *TCR*-β transcription may be mediated by inactivation of Ets-1 DNA-binding activity, possibly by phosphorylation.

A new member of the ets gene family, *elf-1*, has recently been cloned from a human T-cell library and shown to bind to the EBS in both the interleukin-2 (IL-2) enhancer and the human immunodeficiency virus-2 (HIV-2) LTR[16]. The IL-2 enhancer contains two EBS, previously shown to be bound by the NF–AT-1 and NF–IL-2B regulatory transcription factor complexes[16]. Fos and Jun have recently been shown to be present in the NF–AT-1 complex[45]. Does Elf-1 complex with Fos and Jun on the NF–AT-1 site? These sites both bind Elf-1 but not Ets-1 or Ets-2[16]. Other EBS have been identified in lymphoid-specific promoters/enhancers, such as the IgH enhancer, the major histocompatibility complex (MHC) II 1-Aβ enhancer, the HTLV-I LTR, the CD2 enhancer, the CD3 enhancer, the *mb-1* promoter and the terminal deoxynucleotide transferase promoter.

Genes involved in the degradation of the extracellular matrix (ECM), such as those encoding stromelysin-1, collagenase and urokinase plasminogen activator[21], contain EBS in their regulatory flanking sequences. Ets-1 activates expression of rat stromelysin 1 (Ref. 28) and may thus play a role in coordinating events that lead to degradation of the ECM, an important process in metastasis[21].

Ets-1 can act either as an activator of stromelysin 1 in HeLa cells, or as a repressor of TCR-β in quiescent Jurkat cells, depending on promoter/enhancer context and cell type. Analysis of *trans*-activation of different target genes suggests a role for Ets-1 in both T-cell differentiation and in control of ECM degradation.

Trans-activation by Ets

Trans-activation by $p68^{c\text{-}ets\text{-}1}$ from the PEA3 site of the polyomavirus enhancer is cooperative with activation by the transcription factor, AP-1, bound to the adjacent PEA1 site. Cooperativity is not mediated at the level of DNA binding since altering the DNA binding affinity of either the PEA1 or PEA3 sites did not alter the ability of these factors to cooperate in transcription[27]. Over-expression of $p68^{c\text{-}ets\text{-}1}$ resulted in inhibition of *trans*-activation by AP-1 of the oncogene response unit, mutated in its PEA3 site[27]. This suggested that the cooperativity between AP-1 and $p68^{c\text{-}ets\text{-}1}$ is mediated by protein–protein interactions[27]. $p54^{c\text{-}ets\text{-}1}$ and $p58^{c\text{-}ets\text{-}2}$ of chicken also bind to the PEA3 site and effect cooperative *trans*-activation with AP-1[26]. Recent evidence has demonstrated cooperative *trans*-activation by c-Myb and Ets-2 of the *mim-1* (*myb*-induced *my*eloid-*1*) promoter; however, Ets-1 failed to cooperate with c-Myb in *trans*-activation[36]. Cooperativity in *trans*-activation has also been observed between two palindromic EBS in the rat stromelysin 1 gene promoter[28]. A single EBS does not appear to mediate transcriptional activation linked to a minimal promoter[37], and, as in the case of $p75^{c\text{-}myb}$, multiple binding sites are required for this activation function.

Domain-swap experiments involving fusion of deletion mutants of $p68^{c\text{-}ets\text{-}1}$ to the DNA-binding domain of bacterial LexA have identified three *trans*-activation domains, RI, RII and RIII, in the amino-terminal region of $p68^{c\text{-}ets\text{-}1}$ (Ref.

25; Fig. 2). When fused to the LexA DNA-binding domain, RI and RIII can both activate transcription from promoters containing the *lexA* operator[25]. RII cannot *trans*-activate alone, but has a negative regulatory effect on the activity of RI and a positive effect on the activity of RIII. RII, which contains a helix-loop-helix motif[3], is conserved between Ets-1 and Ets-2, suggesting that RII may mediate common control interactions of Ets-1 and Ets-2. Ets-1 and Ets-2 compete with LexA–Ets-1/-2 fusion proteins in *trans*-activation, suggesting that Ets-1 and Ets-2 proteins interact with a common factor.

The *trans*-activation domain of PU-1 has both a Gln-rich region (mapping from amino acids 74–93) and an acidic region (amino acids 1–165), which may act as *trans*-activation domains[12]. Recent work has identified a role for PU-1 in recruiting other transcription factors such as NF-EM5, a B-cell-specific factor, to the 3' enhancer of the mouse immunoglobulin kappa light-chain gene, where binding to the enhancer by NF-EM5 depends upon prior binding of PU-1 to the enhancer[12]. The functional domains involved in these interactions have not been identified.

The serum response factor (SRF) binds to the serum response element (SRE) in the promoters/enhancers of serum-inducible genes, such as *c-fos*. SRF forms a ternary complex with $p62^{TCF}$, which is essential for complex formation on the SRE. $p62^{TCF}$ does not bind DNA alone. Elk-1 is functionally equivalent to $p62^{TCF}$, in that it forms a ternary complex with SRF[10]. Similar results have been obtained with an Elk-related protein, SAP-1[17], which forms a ternary complex with SRF over the SRE of the *c-fos* promoter and, like Elk-1, its DNA-binding activity depends upon interaction with the SRF. Both Elk-1 and SAP-1 are members of the Ets family. Two domains, including the ETS domain of SAP-1, are necessary for ternary complex formation[17].

A new member of the *ets* gene family, GABP-α, forms the α-subunit of the GA-binding protein (GABP), which binds to the purine-rich sequence (CGGAAG/A) of the *ICP4* gene of herpes simplex virus[3,14]. GABP-α mediates DNA binding by the multi-subunit protein but GABP-β is also necessary for binding[13,14]. The amino terminus of the β-subunit contains four 33-amino-acid ankyrin repeats found in the Notch, Lin12, cdc10, NF-κB and Iκ-B proteins[13,14]. These 33-amino-acid repeats are necessary for interaction of GABP-β with GABP-α. The ETS domain of the α-subunit is required for its interaction with the β-subunit.

The transcription factor E4TF1 regulates transcription of the adenovirus-5 *E4* gene and binds to a sequence CGGAAGT, a high-affinity EBS, in the *E4* promoter[38]. E4TF1 is composed of two subunits, p53 and p60, and addition of purified p53 increases the binding affinity of p60 for this site. Neither p53 nor p60 can *trans*-activate alone but together mediate a tenfold increase in transcription mediated by the CGGAAGT site[38]. p60 is a good candidate member of the Ets family.

Ets in transformation

The v-*ets* oncogene is co-transduced with the v-*myb* oncogene in the E26 retrovirus, which expresses a 135 kDa fusion protein, $p135^{\delta gag\text{-}myb\text{-}ets}$ (Refs 1,2). The acute myeloblastosis virus (AMV), which has transduced the v-*myb* oncogene alone, induces myeloid leukaemias[39], whereas E26 induces erythroleukaemias. It was proposed that v-*ets* confers the ability to transform immature erythroid progenitor cells upon the E26 virus. Both v-*ets* and v-*myb* retain a reduced capacity to transform erythroid cells individually, but co-expression results in an increased self-renewal rate of transformed erythroid cells[39]. Clearly, the transforming ability of v-*ets* and v-*myb* is additive, suggesting that v-*myb* and v-*ets* oncogene products may cooperate in transcription[36,39]. This transforming activity is superceded by the transforming potential of *myb–ets* fusion proteins, which generate erythroleukemia with a much-reduced latency and higher frequency[39]. Fusion of the v-*ets* and v-*myb* oncogenes may generate a novel *trans*-activator with an altered target gene specificity or altered interactions with other regulatory factors. More recent work has shown that v-*ets* can also cooperate with v-*erbA* in transforming erythroid progenitor cells[43].

There is a common mechanism that results in the activation of two different *ets* genes in virally induced murine erythroleukemias. Proviral integration by the Friend Murine Leukaemia Virus (Friend MuLV) and the Spleen Focus Forming Virus (SFFV) into flanking sequences activates expression of the *fli-1* and *spi-1* genes respectively.

The human *ets-1* and *ets-2* genes are translocated in certain types of acute leukaemia from chromosomes 11 to 4 and from 21 to 8, respectively, although neither gene spans the chromosome breakpoint[2]. Similarly, the *elk-1* gene maps close to chromosome breakpoints in synovial sarcomas[2]. The significance of these translocations for Ets function is unclear.

The role of *ets* gene family members as oncogenes has been tested in classical transformation experiments in which over-expression of the human c-*ets-1* or c-*ets-2* oncogenes in NIH-3T3 mouse fibroblasts generates transformed foci in low serum conditions[3,41]. These transformed cells grew in soft agar and formed tumours in nude mice. The transforming capacity of the c-*ets-1* proto-oncogene may be associated with its ability to positively autoregulate its own expression[3,41].

Increased binding to the PEA3 element of the polyoma enhancer occurs when fibroblasts are stimulated with serum or phorbol esters or transformed by v-*src*, v-*raf*, polyoma middle T antigen, v-*mos* or c-Ha-*ras* (Ref. 36). This induction of binding activity is independent of protein synthesis inhibitors, suggesting that it is mediated by post-translational modifications, such as phosphorylation or interactions with cofactors. Recent evidence suggests that Raf-1 kinase is directly/indirectly involved in activating Ets-driven promoters following Ras transformation of fibroblasts[42].

An interesting new aspect concerning the role of *ets* in growth control has emerged from studies of Ets-1 expression in human endothelial cells[21]. Ets-1 expression correlates with the proliferation of endothelial cells in normal blood vessel formation, associated with wound healing and normal developmental processes, but also with vascularization of tumours during malignant progression[21]. This evidence, combined with the ability of Ets-1 to regulate the expression of ECM-degrading proteins such as stromelysin and collagenase, suggests an important role for Ets-1 in the metastatic processes associated with late tumour development[21,28].

Perspectives

The number of members in the *ets* gene family is growing rapidly, the encoded proteins of which function as regulators of cellular and viral gene transcription. Some Ets proteins, such as Ets-1 and Ets-2, interact with the same DNA-recognition sites and may function antagonistically, either activating or repressing transcription of the same target gene, depending on promoter context, cell-type or stage of differentiation. Other Ets proteins, such as Elf-1, interact with different, although similar sites, and probably regulate different target genes. How the expression and activity of different Ets proteins are coordinated with respect to cell-type, in differentiation, during the cell cycle and in development, is unclear. The capacity of certain Ets proteins to interact with regulatory cofactors and to be regulated by phosphorylation suggests that Ets proteins may play a role in coordinating changes in gene expression in response to second messengers and extracellular signals. Ets proteins may be modulated differently by interactions with different cofactors. The possible role of these interactions in the regulatory pathways of cell-cycle control and differentiation needs further evaluation.

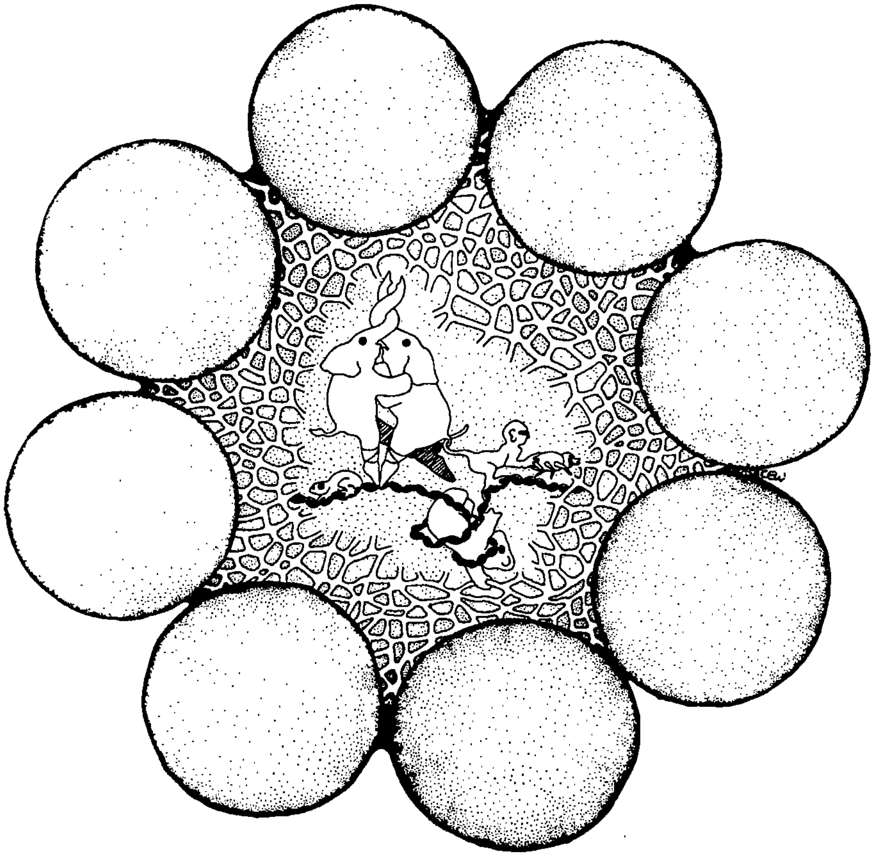
Acknowledgements

The authors would like to thank Drs Y. de Launoit, O. Albagli, P. Martin and S. Saule for critical reading of the manuscript and Dr M. Owen, ICRF, for communication of unpublished data. K. Macleod is financially supported by the Commission of the European Community, D. Leprince and D. Stehelin are funded by the Centre National de la Recherche Scientifique, D. Stehelin is Professor of Molecular Oncology, Institut Pasteur, Lille.

References

Due to *TIBS* policy of short reference lists, the number of references cited in this article has been limited. Some work has therefore unfortunately been left unacknowledged.

1 Leprince, D. *et al.* (1983) *Nature* 306, 395–397
2 Watson, D. K., Ascione, R. and Papas, T. S. (1990) *Crit. Rev. Oncogen.* 1, 409–436
3 Seth, A. and Papas, A. (1990) *Oncogene* 5, 1761–1767
4 Leprince, D. *et al.* (1988) *J. Virol.* 62, 3233–3241
5 Koizumi, S. *et al.* (1990) *Oncogene* 5, 675–681
6 Karim, F. D. *et al.* (1990) *Genes Dev.* 4, 1451–1453
7 Boulukos, K. E. *et al.* (1990) *Genes Dev.* 4, 401–409
8 Reddy, E. S. P. and Rao, V. N. (1991) *Oncogene* 6, 2285–2289
9 Rao, V. N. and Reddy, E. S. P. (1992) *Oncogene* 7, 65–70
10 Hipskind, R. A. *et al.* (1991) *Nature* 354, 531–534
11 Ben-David, Y., Giddens, E. B., Letwin, K. and Bernstein, A. (1991) *Genes Dev.* 5, 908–918
12 Pongubala, J. M. R. *et al.* (1992) *Mol. Cell. Biol.* 12, 368–378
13 Thompson, C. C., Brown, T. A. and McKnight, S. L. (1991) *Science* 253, 762–768
14 LaMarco, K. *et al.* (1991) *Science* 253, 789–791
15 Pribyl, L. J., Watson, D. K., Schulz, R. A. and Papas, T. S. (1991) *Oncogene* 6, 1175–1183
16 Thompson, C. B. *et al.* (1992) *Mol. Cell. Biol.* 12, 1043–1053
17 Dalton, S. and Treisman, R. (1992) *Cell* 68, 597–612
18 Xin, J-H., Cowie, A., Lachance, P. and Hassell, J. A. (1992) *Genes Dev.* 6, 481–496
19 Waterman, M. L., Fischer, W. H. and Jones, K. A. (1991) *Genes Dev.* 5, 656–669
20 Bhat, N. K. *et al.* (1990) *Proc. Natl Acad. Sci. USA* 87, 3723–3727
21 Wernert, N. *et al.* (1992) *Am. J. Pathol.* 140, 119–127
22 Mavrothalassitis, G. J. and Papas, T. S. (1991) *Cell Growth Differ.* 2, 215–224
23 Pognonec, P., Boulukos, K. E. and Ghysdael, J. (1989) *Oncogene* 4, 691–697
24 Oka, T., Rairkar, A. and Chen, J. H. (1991) *Oncogene* 6, 2077–2083
25 Schneikert, J., Lutz, Y. and Wasylyk, B. (1992) *Oncogene* 7, 249–256
26 Fujiwara, S. *et al.* (1988) *Mol. Cell. Biol.* 8, 4700–4706
27 Wasylyk, B. *et al.* (1990) *Nature* 346, 191–193
28 Wasylyk, C., Gutman, A., Nicholson, R. and Wasylyk, B. (1991) *EMBO J.* 10, 1127–1134
29 Fisher, R. J., Mavrothalassitis, G., Kondoh, A. and Papas, T. S. (1991) *Oncogene* 6, 2249–2254
30 Woods, D. B., Ghysdael, J. and Owen, M. J. (1992) *Nucleic Acids Res.* 20, 699–704
31 Gitlin, S. D. *et al.* (1991) *J. Virol.* 65, 5513–5523
32 Gunther, C. V., Nye, J. A., Bryner, R. S. and Graves, B. J. (1990) *Genes Dev.* 4, 667–679
33 Ho, I-C. *et al.* (1990) *Science* 250, 814–818
34 Leprince, D., Crepieux, P. and Stehelin, D. (1992) *Oncogene* 7, 9–17
35 Lim, F., Kraut, N., Frampton, J. and Graf, T. (1992) *EMBO J.* 11, 643–652
36 Dudek, H. *et al.* (1992) *Proc. Natl Acad. Sci. USA* 89, 1291–1295
37 Wasylyk, C., Flores, P., Gutman, A. and Wasylyk, B. (1989) *EMBO J.* 8, 3371–3378
38 Watanabe, H., Wada, T. and Handa, H. (1990) *EMBO J.* 9, 841–847
39 Metz, T. and Graf, T. (1991) *Cell* 66, 95–105
40 Jorcyk, C. L., Watson, D. K., Mavrothalaesitis, G. J. and Papas, T. S. (1991) *Oncogene* 6, 523–532
41 Seth, A., Watson, D. K., Blair, D. G. and Papas, T. S. (1989) *Proc. Natl Acad. Sci. USA* 86, 7833–7837
42 Bruder, J. T., Heidecker, G. and Rapp, U. R. (1992) *Genes Dev.* 6, 545–556
43 Metz, T. and Graf, T. (1992) *Oncogene* 7, 597–605
44 Martin, M. E., Yang, X. Y. and Folk, W. R. (1992) *Mol. Cell. Biol.* 12, 2213–2221
45 Jain, J., McCaffrey, P. G., Valge-Archer, V. E. and Rao, A. (1992) *Nature* 356, 801–804
46 Wasylyk, C., Kerchaert, J. P. and Wasylyk, B. *Genes Dev.* (in press)



After years of arduous research on the regulation of transcription, Dr. Benway suddenly realises the true nature of the nucleus.

Pergamon

Int. J. Biochem. Cell Biol. Vol. 29, No. 12, pp. 1313–1323, 1997
© 1997 Elsevier Science Ltd. All rights reserved
Printed in Great Britain
1357-2725/97 $17.00 + 0.00

PII: S1357-2725(97)00094-0

REVIEW

Transcriptional Regulation by the Sp Family Proteins

LUIGI LANIA,[1]* BARBARA MAJELLO,[1,2] PASQUALE DE LUCA[1,2]

[1]*Department of Genetics, Molecular and General Biology, University of Naples 'Federico II', via Mezzocannone 8, 80 134, Naples, Italy and* [2]*International Institute of Genetics and Biophysics, CNR, Naples, Italy*

Sp1 is one of the very first cellular transcription factors to be identified and cloned in virtue of its binding to a G-rich motif in the SV40 early promoter. Sp1 protein binds to the G-rich sequences present in a variety of cellular and viral promoters and stimulates their transcriptional activity. Recently, a number of other GC and/or GT box-binding factors homologous to Sp1 have been isolated, namely Sp2, Sp3 and Sp4, and the two more distantly related factors, BTEB and BTEB2. The discovery of this family highlights a previously unknown level of complexity of transcriptional regulation of promoters containing GC and/or GT box motifs. This review focuses primarily on strategies aimed to elucidate the transcription properties of the Sp1-like factors and discusses the experimental problems inherent in the attempt to define their respective functions. © 1997 Elsevier Science Ltd. All rights reserved

Keywords: GC-box Activator Repressor Sp1 Sp3 Sp4

Int. J. Biochem. Cell Biol. (1997) **29**, 1313–1323

INTRODUCTION

Our knowledge of the transcriptional regulatory mechanisms governing gene expression is accumulating at an accelerating pace. Such progress stems largely from the combination of biochemical and genetics studies of gene transcription. Initiation of messenger RNA synthesis is the major site for regulation of eukaryotic gene expression. Intense effort has been focused on identifying transcription factors that control mRNA synthesis by RNA polymerase II. These studies recently culminated in the purification and molecular cloning of a set of essential general initiation factors (GTFs) that promote basal transcription by polymerase II. In addition, a large number of sequence-specific DNA-binding transcription activator or repressor proteins have been isolated. It is now widely appreciated that transcription initiation can be broadly divided into three steps: initiation complex assembly, isomerization, and promoter clearance. In the first step RNA polymerase and associated factors bind reversibly to the promoter. In the second step a stretch of promoter DNA becomes unwound and serves as a template for transcription. Finally, the polymerase leaves the promoter to elongate the transcript (Conaway and Conaway, 1994). The efficiency of each step can be subject to regulation by transcription activator or repressor proteins. DNA-bound transcription factors are thought to influence the transcription through protein–protein interaction with components of the general initiation factors resulting in an enhanced recruitment or stabilization of the basal promoter complex to the template. Upon binding their cognate sites, activator proteins stimulate transcription via an activation domain that is functionally distinct, and usually physically separate, from the DNA-binding domain (Latchman, 1991). A subset of activators are expressed in a cell type-specific manner, e.g. Oct-2 in B cells, and regulate the expression of genes specific for

---

*To whom all correspondence should be addressed at University of Naples.

Received 27 February 1997; accepted 16 June 1997.

the given cell type. Other activators, such as Spl, are ubiquitously expressed and are required for the constitutive and inducible expression of a variety of genes. The human transcription factor Spl binds to the so-called GC boxes, and has long been thought to be the single factor acting through these sites (Kadonaga *et al.*, 1987). However, this view has changed following the isolation of cDNAs encoding Spl-related proteins, indicating the presence of a novel Spl multigene family. To date three Spl-related proteins, which have been called Sp2, Sp3 and Sp4 have been described (Hagen *et al.*, 1992; Kingsley and Winoto, 1992). They contain zinc fingers as well as glutamine and serine/threonine rich amino acid stretches similar to those of Spl. The DNA binding domains of the Sp2, Sp3 and Sp4 proteins are highly conserved and they recognize GC boxes with specificity and affinity similar to that of Spl. Two more distantly related factors, BTEB and BTEB2, have been isolated by virtue of their ability to bind a GC-box (Imataka *et al.*, 1992; Sogawa *et al.*, 1993). Clearly, the existence of proteins with DNA-binding specificity similar to Spl indicates that gene regulation by Spl is more complicated than previously assumed.

ISOLATION OF THE SP1-FAMILY MEMBERS

The human transcription factor Spl was originally identified as a factor able to bind and activate the SV40 early promoter. The Spl protein binds selectively to a GC-rich decanucleotide sequence known as the "GC box", which is present as six tandem copies in the SV40 early promoter (Kadonaga *et al.*, 1987). Subsequent studies revealed that a variety of cellular and viral promoters contain GC boxes and can be activated by Spl *in vitro*; these GC boxes are often found near binding sites for other transcription factors, suggesting that these factors may act in conjunction with each other to modulate transcription. Cloning of the Spl cDNA from HeLa cell RNA showed that it contained three putative zinc fingers of the Cys_2His_2 type (Kadonaga *et al.*, 1987). Functional analysis have delineated portions of Spl required for activation, DNA-binding and multimerization, respectively (Courey and Tjian, 1988; Courey *et al.*, 1989; Pascal and Tjian, 1991). The Spl trans-activation domain has been subdivided into two subdomains (A and B) with regions rich in serine/threonine and glutamine residues, and a region close to the zinc fingers largely composed of charged amino acids, which synergistically stimulate transcription. The Spl protein has also been shown to be modified with O-linked oligosaccharides *in vivo* and phosphorylated by a DNA-dependent protein kinase (Jackson and Tijan, 1988; Jackson *et al.*, 1990). Thus Spl may be a target of intracellular signalling, although it is not yet clear whether these post-translation modifications play any role in regulating or modifying the Spl-mediated transcription. Moreover, it is uncertain whether Spl-mediated transcription is actively regulated in mammalian cells or is simply determined by the amount of Spl protein within the nucleus. The mouse Spl gene was shown to be ubiquitously expressed but different cell types and different stages of development show up to 100-fold differences in the levels of Spl mRNA (Saffer *et al.*, 1990, 1991). Based on these considerations, it is likely that Spl acts in combinatorial manner with other transcription factors, which may have more pronounced temporally or spatially restricted expression patterns.

Several proteins with homology to Spl have been identified in a number of different systems (Fig. 1). A rat protein designed BTEB was cloned by virtue of binding the BTE box, a GC box sequence in the promoter of the rat P4501A1 gene (Imataka *et al.*, 1992). Both Spl and BTEB affect expression of P4501A1 through the BTE sequence. Using the rat BTEB cDNA as a probe, a related cDNA designated BTEB2 was isolated from human placenta library (Sogawa *et al.*, 1993). This gene was found to be expressed specifically in testes and placenta, but no known function has yet been described. The two Spl-related proteins called Sp2 and Sp3 were both identified for their binding to a GT box motif in the T-cell antigen receptor (TCR) α promoter (Kingsley and Winoto, 1992). These genes were cloned based on the sequence similarity of their zinc fingers to those of Spl. Sp2 protein was found to bind the GT box motif weakly, whereas Sp3 bound this sequences with high affinity comparable to that of Spl. Two human transcription factors, designated originally SPR-1 and SPR-2 (for "Spl related proteins"), were isolated because of their ability to bind a GT box motif in the rabbit uteroglobin promoter (Hagen *et al.*, 1992). In an attempt to identify transcription factors re-

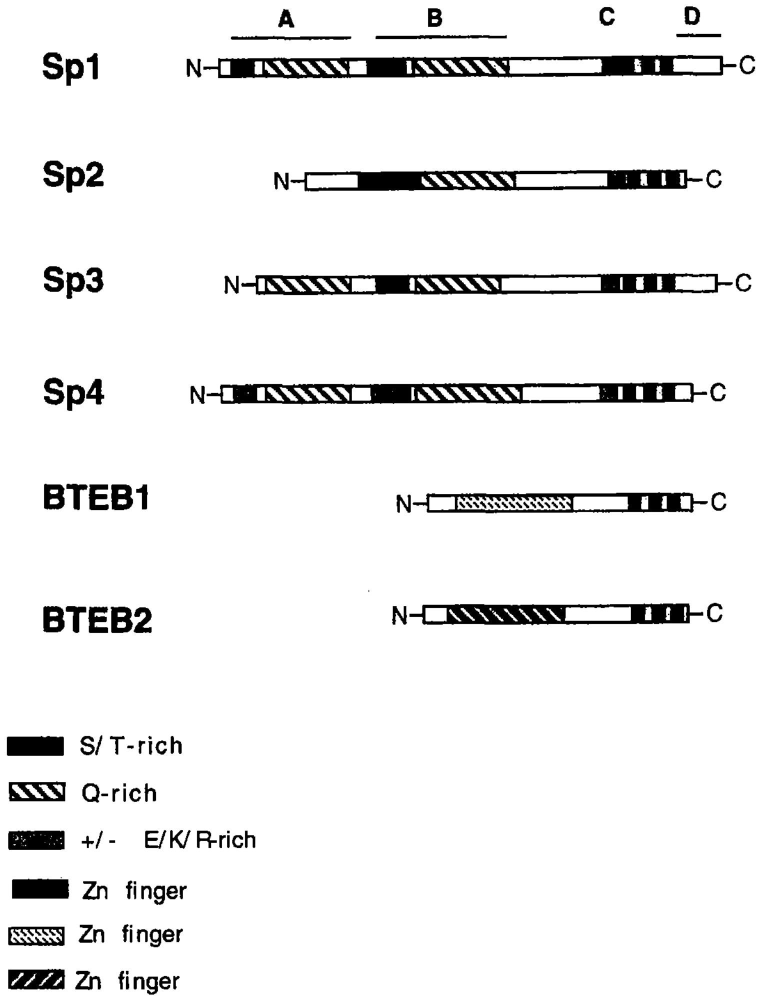


Fig. 1. Schematic representation of the functional domains of Sp1 and comparison with those present in the related Sp1-like proteins.

sponsible for the relatively high expression of uteroglobin in particular cell types, the human endometrial cells (Ishikawa cell line) and the lung cell line NCI-H441, an expression library made from Ishikawa cell mRNA was screened with a concatemer of the GT box binding site and SPR-1 and a related SPR-2 cDNA were cloned (Hagen *et al.*, 1992). SPR-2 is identical to Sp3, and SPR-1 has been renamed Sp4. The zinc finger regions of these proteins and of Sp1 bind to both GT and GC box motifs with similar affinities. The discovery of a family of transcription factors in vertebrates with homologous zinc finger amino acid sequence (Fig. 2) and similar DNA binding properties raises the intriguing possibility that genes once thought to be regulated by Sp1 might be controlled *in vivo* by protein(s) other than Sp1. Another homologue of Sp1 is the *buttonhead* (*btd*) gene of Drosophila, which is expressed in the head anlagen of early Drosophila embryos (Wimmer *et al.*, 1993). The *btd* gene plays an important role during Drosophila development, with btd mutant flies having abnormalities in head morphogenesis. A transgene expressing the human Sp1 gene was able to partially rescue the phenotype of btd mutant flies (Wimmer *et al.*, 1993), demonstrating at least some degree of functional conservation between the Drosophila and vertebrate proteins.

STRUCTURE AND EXPRESSION OF Sp-LIKE PROTEINS

Inspection of the Sp1-like cDNAs isolated in different laboratories has revealed the presence of a highly conserved DNA binding domain located near the C-terminus of the proteins. As represented in Fig. 1, each protein

```
Sp1     GKKKQHICHIQGCGKVYGKTSHLRAHLRWHTGER
Sp2     Q-----V---PD---TFR---L----V-L-----
Sp3     ----------P-------------------S---         finger 1
Sp4     ----------E-----------------------
BTEB    ASE-R-K-PYS--------S---K--Y-V-----
BTEB2   E-RRI-Y-DYP--T---T-S---K----T----K

Sp1         PFMCTWSYCGKRFTRSDELQRHKRTHTGEK
Sp2         --V-N-FF---------------A-----D-
Sp3         --V-N-M----------------R-------        finger 2
Sp4         --I-N-MF---------------R-------
BTEB        --P---PD-L---S-----T--Y-------
BTEB2       -YK---EG-DW--A-----T--Y-K---A-

Sp1         KFACPECPKRFMRSDHLSKHIKTHQNKK
Sp2         R-E-AQ-Q---------T--Y---LVT-
Sp3         --V----S---------A----------           finger 3
Sp4         R-V----S------------V-------
BTEB        Q-R--L-E---------T--ARR-TDFH
BTEB2       P-Q-GV-NRS-S-----AL-M-R---
```

Fig. 2. Comparison of the amino acid sequences of the zinc finger region of Sp1 and related proteins as indicated. Dashes represent identical residues.

contains three zinc fingers with structure of Cys$_2$–His$_2$. Homology to Sp1 at the zinc finger domain varies from 59% for BTEB2 to 90% for Sp3 and Sp4. Analysis of the DNA-binding activities revealed that Sp3, Sp4, BTEB and BTEB2 proteins recognize GC and/or GT-box motifs with specificity and affinity very similar to that of Sp1 (Hagen *et al.*, 1992; Imataka *et al.*, 1992; Kingsley and Winoto, 1992; Sogawa *et al.*, 1993; Hagen *et al.*, 1994). Contrary to these factors, Sp2 which binds to the GT box in the promoter of a T-cell receptor gene, seems to have divergent nucleotide recognition properties (Kingsley and Winoto, 1992). The predicted Sp3 and Sp4 protein sequences have extensive homology to Sp1 throughout the entire open reading frame, whereas homology outside the zinc finger domain of Sp2 is much more limited.

Both BTEB and BTEB2 do not show any homology to Sp1 outside the DNA-binding domain. BTEB has an acidic region and several transcription factors such as GAL4 are equipped with an acidic domain as effector (Mitchell and Tjian, 1989). BTEB has been tested for its ability to regulate gene transcription and both Sp1 and BTEB activated the expression of promoters with repeated GC box sequences such as the SV40 early promoters and the HIV-1 LTR promoters (Imataka *et al.*, 1993). Interestingly, BTEB represses the activity of the promoter containing a single GC box that instead can be stimulated by Sp1 (Imataka *et al.*, 1992). However, when the GC-box was repeated five times, BTEB turned out to be an activator. Therefore, it appears that BTEB could exert a different effect on the transcription of those genes with a different number and arrangement of GC-box sequences in the promoter region. Although BTEB mRNA is expressed in various mammalian tissues and cell lines, the presence of the BTEB protein is confined to brain and some cell lines, such as Neuro2A cells. It has been shown that BTEB mRNA is differentially translated depending upon the cellular context (Imataka *et al.*, 1994). Finally, the human BTEB cDNA has been cloned and mapped to the chromosome 9q13 (Ohe *et al.*, 1993). BTEB2 is characterized by the presence of a Proline-rich domain, which is known to be one of the motifs responsible for transactivation (Latchman, 1991). Accordingly, the BTEB2 proline-rich region fused to GAL4 DNA-binding domain could enhance transcription of a reporter bearing the GAL4 DNA-binding sites (Sogawa *et al.*, 1993).

Clearly, Sp1, Sp3 and Sp4 are the Sp-members with the most close homology. All three proteins contain highly conserved DNA binding domains consisting of three zinc finger motifs at the carboxyl terminus. Sp3 and Sp4 also contain several putative functional domains homologous to those of Sp1. These include: (i) the amino-terminal glutamine- and serine/threonine-rich regions similar to the A and B transactivation domains of Sp1; (ii) the highly charged domain immediately 5′ of the zinc fingers; (iii) a carboxy-terminal domain similar to the D transactivation domain of Sp1 present in Sp3. However, the expression pattern of the three proteins is quite different. Both Sp1 and Sp3 proteins are ubiquitously expressed at a high level in many mammalian cell lines (Fig. 3), whereas Sp4 expression appears to be restricted to certain cell types of the brain. Finally, the chromosomal localization of the three genes has been reported. Sp1 has been mapped on chromosome 12q13 (Gaynor *et al.*, 1993; Matera and Ward, 1993), Sp3 is located at 2q31 (Kalff-Suske *et al.*, 1996) and Sp4 on chromosome 7p15 (Kalff-Suske *et al.*, 1995).

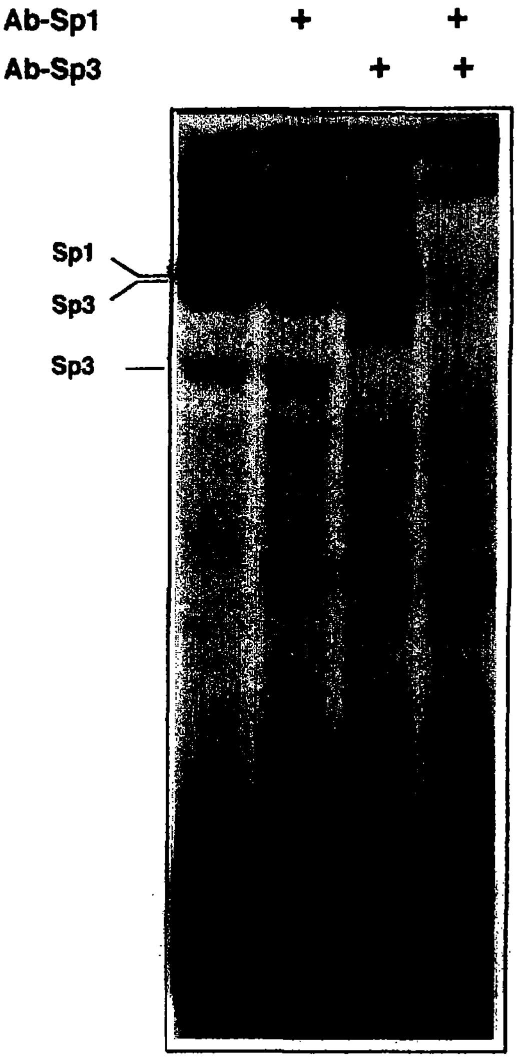


Fig. 3. Gel mobility shift analysis using as probe a labelled oligonucleotide spanning the three Sp1-binding sequences present in the promoter of the HIV-1 LTR, and a nuclear extract from HeLa cells. The presence of specific antibodies in the binding reaction is indicated above each line. Specific complexes for Sp1 and Sp3 are indicated on the left.

The high degree of structural conservation between Sp1, Sp3 and Sp4 suggested that Sp3 and Sp4 may regulate gene transcription in a manner similar to that of Sp1. In the following sections we will describe the functional properties of these factors in direct comparison with those of Sp1.

Sp1 IS REQUIRED FOR EARLY DEVELOPMENT

A large number of observations imply a wider role of the Sp1 protein in the transcriptional regulation of tissue-specific and ubiquitous genes. Moreover, Sp1 has been implicated in the control of cell cycle-regulated genes such as thymidine kinase, B-myb and dihydrofolate reductase (Karlseder *et al.*, 1996, Lin *et al.*, 1996, Zwicker *et al.*, 1996). It has been suggested that the activity of Sp1 may be regulated by members of the retinoblastoma family proteins pRB and p107 (Kim *et al.*, 1992; Udvadia *et al.*, 1993, 1995; Chen *et al.*, 1994; Datta *et al.*, 1995) and by G1-specific cyclins (Shao and Robbins, 1995). In addition, Sp1-binding sites appear to play a critical role for the maintenance of the methylation-free CpG island (Brandeis *et al.*, 1994, Macleod *et al.*, 1994). Thus, it appears that Sp1 plays an important role in controlling putative cell-cycle regulated genes and it is required to prevent methylation of CpG islands. Despite a great deal of information on the molecular properties of Sp1, little is known about its biological function *in vivo*. Very recently Marin and collaborators (Marin *et al.*, 1997) reported the inactivation of the mouse Sp1 gene and the characterization of the phenotype of the Sp1-null mice. They found that the Sp1-/- embryos are retarded in development, survive until day 9.5 of gestation and show a broad range of phenotypic abnormalities. The broad range of abnormalities suggests that Sp1-deficiency causes a general cellular defect that precludes normal development and survival. Due to the large number of putative Sp1 target genes it was not surprising that inactivation of Sp1 is not compatible with cellular growth and differentiation. However, Marin *et al.* (1997) also looked at the expression of putative Sp1 genes at the "housekeeping" and cell cycle-regulated TK, DHFR, AFRT and

HPRT, and some tissue-restricted as ApoAl and LDLR genes no differences were found. Moreover, they also analyzed the CpG islands and found that the islands remain methylation free. The only clear differences in the Spl embryos was found at the level of MeCP2 expression. This protein binds the methylated DNA and it appears to be required for the maintenance of the differentiated cells (Tate *et al.*, 1996). Marin *et al.* (1997) suggested that Spl is dispensable for growth and differentiation of primitive cells and suggest that Spl is likely to play a pivotal role in the maintenance of differentiated cells likely through transcriptional regulation of genes such as MeCP2. It is safe to assume that other members of the Spl family, like Sp3 and or Sp4, may compensate at least in part for the loss of Spl activity in the transcriptional regulation of cell-cycle regulated and tissue restricted genes. Despite this potential compensatory function, the data reported by Marin *et al.* clearly demonstrated that Spl-dependent gene activation is strictly required for mouse development.

Sp4 IS A TISSUE-SPECIFIC TRANSCRIPTIONAL ACTIVATOR

The first indication that Sp4 functions as an activator much as Spl, came from transfection experiments into mammalian cell lines. According to the notion whereby Sp4 is a transcriptional activator, it was found that GAL4 fusion proteins, in which the two glutamine-rich domains were fused to the GAL4 DNA-binding domain, exert the potential for transactivation of appropriate reporters, like those present in Spl (Majello *et al.*, 1994; Dennig *et al.*, 1995; Hagen *et al.*, 1995). A detailed functional analysis of the transcriptional properties of Sp4 has been recently reported (Hagen *et al.*, 1995). These studies demonstrated that Sp4 is an activator protein like Spl. It has been shown that Sp4 is able to trans-activate viral and synthetic promoters bearing GC or GT box motifs both in mammalian cells and in Drosophila SL2 cells which are devoid of endogenous Sp-like activity. Moreover, transfection experiments in Drosophila SL2 cells revealed that in contrast to Spl, Sp4 is not able to act synergistically through adjacent binding sites. The region of Spl protein required for synergistic activation has been mapped (Pascal and Tjian, 1991). Sequence comparison (Fig. 1) indicates that Sp4 lacks the D-domain present in Spl. It has been shown that the D-domain is indispensable for Spl-mediated synergistic transactivation. The absence of the D-domain in Sp4 may thus account for the lack of Sp4-mediated synergistic activation. Interestingly, Sp4-mediated activation is strongly enhanced in the presence of Spl mutant lacking the DNA binding domain. This finding could imply that the glutamine-rich domains of Sp4 and those of Spl are functionally related to each other (Hagen *et al.*, 1995).

Northern analysis clearly indicates that Sp4 expression *in vivo* is restricted to certain cell types of the brain (Hagen *et al.*, 1992). Thus, Sp4 may play an important role in governing expression of certain genes in these cells. This hypothesis has recently received experimental support by the study of the mouse sp4 gene during development (Supp *et al.*, 1996). This study reported the isolation and mutation by gene targeting of the mouse Sp4 gene. The Sp4 gene was cloned from mouse because it is located within a region of chromosome 12 deleted in the mutant mouse legless (lgl). Sp4 gene has an interesting pattern of expression during embryonic development. *In situ* hybridizations showed that Sp4 is highly expressed in developing mouse brain. In early embryos, it is expressed in the posterior neuropore; layer, high expression levels were seen throughout the entire CNS, including the hypothalamic region and in the nasal epithelium and vomeronasal organs. Lower levels of expression have been detected in several mesenchymal tissues such as the liver, testes and teeth. In adult mice, the gene is expressed more widely, with the highest levels seen in the brain. Analysis of Sp4-null mice indicated that this gene plays a role for normal growth and male fertility. Two-thirds of Sp4 homozygous mutants die within the first few days after birth, and those that survive are smaller than their wild-type litter-mates. The cause of the early death remains undetermined. The only other detectable defect of these mutants is the lack of sexual activity in Sp4-null males, although structural abnormalities of the male reproductive tract were not detected. The testes, epididymis, vas deferens and prostates of the mutants were indistinguishable from wild type. Moreover, histological sections of the mutant testes revealed the presence of mature sperm. Therefore, a possible interpretation of these findings could be that the mutation has caused a behavioural defect that affects male reproduction.

Sp3 IS A BIFUNCTIONAL TRANSCRIPTION REGULATOR

Both Spl and Sp3 proteins are ubiquitously expressed in many mammalian cell lines, and as reported in Fig. 3, both proteins are expressed at comparable levels. Transfection experiments into mammalian cell lines indicated that Sp3 was not able to activate various Spl-responsive promoters (Hagen *et al.*, 1994; Majello *et al.*, 1994, 1995; Dennig *et al.*, 1995). However, a severe limitation of these experiments is due to the presence, in all the mammalian cell lines tested, of endogenous Spl and Sp3 proteins which could obscure and/or complicate the interpretation of these transfections. The *Drosophila* SL2 cell line lacks endogenous Sp factors, therefore it has been instrumental in analyzing the Sp3-mediated transcription. It has been originally found that the Sp3 failed to activate Spl-responsive promoters such as HIV-1 and HTLV-1 in *Drosophila* SL2 cells (Hagen *et al.*, 1994; Majello *et al.*, 1994). Moreover, cotransfection experiments in *Drosophila* SL2 cells indicated that enforced expression of Sp3 represses Spl-mediated activation of several viral and cellular promoters. A likely interpretation of these results is that Sp3 lacks transactivation potential and it can suppress the Spl activity simply because it binds to, and thus competes for, the same DNA binding sites. Therefore, Sp3-mediated transcriptional repression can result from competition for the same DNA-binding sites or steric hindrance between repressors and positively acting transcription factors, such as Spl. Because this type of repression results from displacement of transcription factors from the DNA, Sp3 might not necessarily possess an active repression function. Alternatively, Sp3 may actively repress Spl-dependent transcription due to the presence of a putative repressor domain. Such hypothesis is supported by the finding that a fusion protein consisting of the yeast GAL4 DNA binding domain and the N-terminal part of Sp3 containing the glutamine-rich domains did not activate GAL4-responsive promoters in mammalian cell lines. Moreover, it was found that Tat-mediated activation of several HIV-1 LTR reporters, containing five Gal4 DNA-binding sites inserted at different distances from the transcription start site, was abrogated by co-expression of the chimeric GAL4-Sp3 protein (Majello *et al.*, 1994). Distance-independent repression observed with the GAL4-SP3 fusion protein rules out the possibility that Sp3 is simply sterically blocking the promoter. Repression at distance supports a mechanism requiring protein–protein interactions between distantly bound Sp3 protein and proximal factors, such as TAFs, looping out intervening DNA (Licht *et al.*, 1993, Hanna-Rose and Hansen, 1996). This conclusion has received further support by a study describing *in vivo* transfections, in which the non-finger region of Sp3 and a defined activating domain connected to different heterologous DNA-binding domains were both targeted to a promoter (De Luca *et al.*, 1996). These results suggest that Sp3 may act as a transcriptional repressor of RNA polymerase II promoters by protein–protein interaction with components of the general transcription complex (De Luca *et al.*, 1996).

Based on the above mentioned results, it appears that Sp3 is a DNA-binding dependent repressor. However, in a recent report Sp3 has been described as a weak activator of several cellular promoters in *Drosophila* SL2 cells and stimulate transcription by functional interaction with the retinoblastoma protein (Udvadia *et al.*, 1995). Moreover, using *Drosophila* SL2 cells, Birnbaum *et al.* (1995) examined the ability of Sp3 to repress Spl-driven transcription from several eukaryotic promoters that contain GC boxes. These investigators found that Sp3 repressed Spl activation with the dihydrofolate reductase promoter, which contains multiple functional GC-boxes. In contrast, Sp3 was not capable of affecting Spl activation of the histone H4 or the thymidine kinase promoters, both cell cycle-controlled promoters containing single GC-boxes. These studies suggest that Sp3 abrogation of Spl-driven transcription is promoter-specific.

Two very recent independent reports demonstrated that Sp3 has the potential to activate transcription in both mammalian and *Drosophila* cells (Dennig *et al.*, 1996; Majello *et al.*, 1997). Both studies have used gene fusion experiments to dissect the functional domains of Sp3. It has been demonstrated that Sp3 is a bifunctional protein containing independent modular repressor and activator domains. The activation potential of Sp3 is distributed over two glutamine-rich N-terminal regions. Both glutamine-rich domains of Sp3 can stimulate transcription as efficiently as the corresponding Spl glutamate domains. Thus, both proteins have an N-terminus glutamine-rich region that functions as a transferable activation domain. The Sp3 transcriptional repressor activity has been mapped in a small amino acid region located at the 5′ of the zinc finger DNA-binding domain, and this domain appears to be sufficient to confer repressor function when fused to a heterologous DNA-binding domain (Majello *et al.*, 1997). Finally, it has been shown that all chimeric GAL4 proteins containing the Sp3 inhibitory/repressor domain function as transcriptional repressors when bound to a promoter bearing multiple DNA-binding sites. These results suggest that the repressor domain is able to mask the glutamine-rich activating domain. Accordingly, GAL4 fusions lacking the repressor/inhibitory domain function as transcriptional activators. The major finding reported in these studies relay in the demonstration that Sp3 protein has the potential to activate transcription due to the presence of the glutamine-rich domains located at the N-terminus (Dennig *et al.*, 1996; Majello *et al.*, 1997). These domains are very similar to the glutamine-rich activating domains present in Spl. *In vivo* and *in vitro* cooperation between Spl and TFIID has been well documented. It has been demonstrated that the glutamine-rich activation domain of Spl can interact with the TBP-associated factor dTAF110 and TBP itself (Emili *et al.*, 1994; Gill *et al.*, 1994). Consistent with the structural similarity of the Spl and Sp3 activating domains, a functional cooperation between the Sp3 protein and components of the basal transcription complex such as TAF110 and TBP has been reported (Dennig *et al.*, 1996; Majello *et al.*, 1997).

In summary, from these studies it appears that Sp3 is a dual-function regulator and the context of cognate DNA-binding sites in a promoter appears to be one of the elements that determines the strength of the Sp3-mediated repression.

Whatever the molecular mechanism responsible for the function of Sp3 may be, it is also clear that its repression ability is dependant upon the cellular context. Sjottem *et al.* (1996) reported that Sp3 had no stimulatory effect on transcription from the LTR of the HERV-H family human endogenous retrovirus-like elements in HeLa cells. However, in the embryonal carcinoma cell line NTera2-D1, Sp3 stimulated the transcriptional activity of the LTR promoters Transcriptional down-regulation of α2-integrin and E-cadherin occurs in human mammary epithelial cells overexpressing Erb-B2 oncogene. It has been shown that Sp3 DNA-binding activity to the α2-integrin promoter is not altered in Erb-B2-overexpressing cells, whereas the Spl DNA-binding activity is reduced. Thus, it is plausible that Sp3 is directly responsible for the downregulation of α2-integrin gene expression in Erb-B2 overexpressing cells (Ye *et al.*, 1996). An additional example of the cell-type specific role of Sp3 has been observed upon *in vitro* induced differentiation of primary human keratinocytes. In primary keratinocytes, Sp3 protein levels exceed those of Spl and this ratio became inverted after *in vitro* differentiation (Apt *et al.*, 1996). Prowse *et al.* (1997) reported incidences that Sp3 is likely involved in the transcriptional induction of the cyclin-dependent kinase inhibitor p21 during terminal differentiation of keratinocytes. Moreover, a role for Sp3 acting as a repressor has been proposed for the elastin gene transcription. The insulin-like growth factor-I (IGF-I) increases elastin gene transcription in aortic smooth muscle cells and this up-regulation is accompanied by the loss of protein binding to the promoter (Conn *et al.*, 1996), and it has been suggested that IGF-I-mediated increase in elastin transcription occurs via a mechanism of derepression involving the abrogation of Sp3 binding to the GC-box present in the proximal promoter. Finally, Grekova *et al.* (1996) found that Sp3 gene transcription is suppressed in peripheral blood mononuclear cells from most Multiple Sclerosis (MS) patients. However, at this stage it is uncertain whether the lack of Sp3 expression is directly involved in the MS phenotype or if it simply reflects genetic heterogeneity.

CONCLUSION AND FUTURE PERSPECTIVE

Despite a great deal of recent progress in the understanding of the transcriptional control of gene expression by different members of the Sp-like factors (Table 1), important questions remain unanswered. A key point is the elucidation of the role of specific Sp members in the control of gene transcription of a

Table 1. Sp factors

Factor	mRNA	Protein	Expression	Mapping	Function	Association	"Null" mice
Sp1	8.2 kb	90 kDa	Ubiquitous	12q13	Activator	TBP; TAF110	Lethal; required for early development
Sp3	5 kb	97 kDa 60 kDa	Ubiquitous	2q31	Repressor Activator	TAF110 TBP;	N.D.
Sp4	7 kb	82 kDa	Brain restricted	7p15	Activator	N.D.	Required for male fertility

N.D. not determined

given promoter. The presence of a set of factors that interact with common binding sites raises the cogent question of how a promoter is regulated by different transcription factors with a similar DNA binding specificity. The molecular anatomy of several promoters reveals the presence of G-rich motifs often found adjacent to binding sites for other transcription factors, suggesting that these factors may act in conjunction with each other to modulate transcription. A physical and/or functional interaction between Sp1 and other transcription factors has been documented. A cooperative interaction between Sp1 and NF-$_{\kappa}$B is required for HIV-1 enhancer in Jurkat T cells (Perkins *et al.*, 1993). Such cooperative interaction has not been observed with the Sp4 protein, although both Sp1 and Sp4 proteins activate basal and Tat-mediated expression of HIV-1 promoter (Majello *et al.*, 1994). p53 and Sp1 appear to interact and cooperate in Tumor Necrosis Factor-induced transcriptional activation of HIV-1 promoter (Gualberto and Baldwin, 1995). Sp1 and E2F-1 factors appear to functionally and physically interact and regulate transcription of the hamster dihydrofolate reductase and the mouse thymidine kinase promoters (Karlseder *et al.*, 1996; Lin *et al.*, 1996). It will be of great interest to determine whether such functional cooperations are specific for Sp1, since they could also provide a functional assay to distinguish between the various Sp-related factors.

A further level of complexity in the regulation of genes containing Sp-binding sites in their regulatory region has been heightened by experiments showing that Sp1-dependent transcription is influenced by specific cell-cycle regulator proteins. It has been reported that RB and p107, two members of the retinoblastoma protein family, activate or repress Sp1-dependent transcription in a cell-type-dependent manner (Kim *et al.*, 1992; Udvadia *et al.*, 1993, 1995; Chen *et al.*, 1994; Datta *et al.*, 1995). Furthermore, it has been suggested that RB indirectly stimulates Sp1 transactivation by liberating Sp1 from a putative Sp1 negative regulator factor Sp1-I (Chen *et al.*, 1994; Murata *et al.*, 1994). Thus, an emerging theme in transcriptional control is that regulation can result from functional interactions between transcription factors. Functional interactions between Sp proteins and the cell-cycle regulators RB and p107 indicate that such cooperation may occur on promoters containing Sp-binding sites. Interestingly, both RB and p107 proteins are intimately involved in the control of cell-cycle regulated transcription (Weinberg, 1995) and several cell cycle-regulated genes, including the c-myc, dhfr, cyclinE, cdk2 and E2F-1, contain Sp-binding sites in their regulatory regions (Azizkhan *et al.*, 1993). The Sp1/Sp3-pRB family interactions may represent a fine regulation of transcription by utilizing a multiplicity of possible interactions.

Acknowledgements—We thank Dr U. Di Porzio for critical discussions and Dr Geoff Nette for critical reading of the manuscript. The authors gratefully acknowledge support by grants from the Italian Association for Cancer Research (AIRC) and from the AIDS Program, Ministero della Sanità.

REFERENCES

Apt D., Watts R. M., Suske G. and Bernard H-U. (1996) High Sp1/Sp3 ratios in epithelial cells during epithelial differentiation and cellular transformation correlate with the activation of the HPV-16 promoter. *Virology* **224,** 281–291.

Azizkhan J. C., Jensen D. E., Pierce A. J. and Wade M. (1993) Transcription from TATA-less promoters: didhydrofolate reductase as a model. *Crit. Rev. Eukaryotic Gene Expression* **3,** 229–254.

Birnbaum M. J., van Wijnen A. J., Odgren P. R., Last T. J., Suske G., Stein G. S. and Stein J. L. (1995) Sp1 trans-activation of cell cycle regulated promoters is selectively repressed by Sp3. *Biochemistry* **34,** 16503–16508.

Brandeis M. J., Frank D., Keshet I., Siegfried Z., Mendelsohn M., Nemes A., Temper V., Razin A. and Cedar H. (1994) Spl elements protect a CpG island from de novo methylation. *Nature* **371**, 435–438.

Chen L. I., Nishinaka T., Kwan K., Kitabayashi I., Yokoyama K., Fu Y-H. F., Grunwald S. and Chiu R. (1994) The retinoblastoma gene product RB stimulates Spl-mediated transcription by liberating Spl from a negative regulator. *Mol. Cell. Biol.* **14**, 4380–4389.

Conaway R. C. and Conaway J. W. (1994) *Transcription mechanisms and regulation.* Raven Press Ltd, New York.

Conn K. J., Rich C. B., Jensen D. E., Fontanilla M. R., Bashir M. M., Rosenbloom J. and Foster J. A. (1996) Insulin-like growth factor-I regulates transcription of the elastin gene through a putative retinoblastoma control element. *J. Biol. Chem.* **271**, 28853–28860.

Courey A. J., Holtzaman D. A., Jackson S. P. and Tjian R. (1989) Synergistic activation by the glutamine-rich domains of human transcription factor Spl. *Cell* **59**, 827–836.

Courey A. J. and Tjian R. (1988) Analysis of Spl *in vivo* reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55**, 887–898.

Datta P. K., Raychaudhuri P. and Bagchi S. (1995) Association of p107 with Spl: genetically separable regions of p107 are involved in regulation of E2F- and Spl-dependent transcription. *Mol. Cell. Biol.* **15**, 5444–5452.

De Luca P., Majello B. and Lania L. (1996) Sp3 represses transcription when tethered to promoter DNA or targeted to promoter proximal RNA. *J. Biol. Chem.* **271**, 8533–8536.

Dennig J., Beato M. and Suske G. (1996) An inhibitor domain in Sp3 regulates its glutamine-rich activation domains. *EMBO J.* **15**, 5659–5667.

Dennig J., Hagen G., Beato M. and Suske G. (1995) Members of the Sp transcription factor family transcription from the uteroglobin promoter. *J. Biol. Chem.* **270**, 12737–12744.

Emili A., Greenblatt J. and Ingles C. J. (1994) Species-specific interaction of the glutamine-rich activation domains of Spl with the TATA box-binding protein. *Mol. Cell. Biol.* **14**, 1582–1593.

Gaynor R. B., Shieh B. H., Klisak I., Sparkes R. S. and Lusis A. J. (1993) Lacalization of the transcription factor SP1 gene to human chromosome 12q12-q13.2. *Cytogenet. Cell Genet.* **64**, 210–212.

Gill G., Pascal E., Tseng Z. H. and Tjian R. (1994) A glutamine-rich hydrophobic patch in transcription factor Spl contacts the dTAFII110 component to the Drosophila TFIID complex and mediates transcriptional activation. *Proc. Natl Acad. Sci. U.S.A.* **91**, 192–196.

Grekova M. C., Robinson E. D., Faerber M. A., Katz P., McFarland H. F. and Richert J. R. (1996) Deficient expression in multiple sclerosis of the inhibitory transcription factor Sp3 in mononuclear blood cells. *Ann. Neurol.* **40**, 108–112.

Gualberto A. and Baldwin A. S. (1995) p53 and Spl interact and cooperate in the tumor necrosis factor-induced transcriptional activation of the HIV-1 long terminal repeat. *J. Biol. Chem.* **270**, 19680–19683.

Hagen G., Dennig J., Preiss A., Beato M. and Suske G. (1995) Functional analyses of the transcription factor Sp4 properties distinct from Spl and Sp3. *J. Biol. Chem.* **270**, 24989–24994.

Hagen G., Muller S., Beato M. and Suske G. (1992) Cloning by recognition site screening of two novel GT box binding proteins: a family of Spl related genes. *Nucleic Acids Res.* **20**, 5519–5525.

Hagen G., Muller S., Beato M. and Suske G. (1994) Spl-mediated transcriptional activation is repressed by Sp3. *The Embo J.* **13**, 3843–3851.

Hanna-Rose W. and Hansen U. (1996) Active repression mechanisms of eukryotic transcription repressors. *TIG* **12**, 229–234.

Imataka H., Mizuno A., Fujii-Kuriyama Y. and Hayami M. (1993) Activation of the human immunodeficiency virus type 1 long terminal repeat by BTEB, a GC box-binding transcription factor. *Aids Res. Hum. Retroviruses* **9**, 825–831.

Imataka H., Nakayama K., Yasumoto K., Mizuno A., Fujii-Kuriyama Y. and Hayami M. (1994) Cell-specific translational control of transcription factor BTEB expression. The role of an upstream AUG in the 5'-untranslated region. *J. Biol. Chem.* **269**, .

Imataka H., Sogawa K., Sasano K., Kobayashi A., Hayami M. and Fujii-Kuriyama Y. (1992) Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-45091A1 gene. *The EMBO J.* **11**, 3663–3671.

Jackson P. S., MacDonald J. J. and Tjian R. (1990) GC box binding induces phosphorylation of Spl by a DNA-dependent protein kinase. *Cell* **63**, 155–165.

Jackson S. P. and Tijan R. (1988) O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* **55**, 125–133.

Kadonaga J. T., Carner K. R., Marsiarz F. R. and Tjian R. (1987) Isolation of cDNA encoding transcription factor Spl and functional analysis of the DNA binding domain. *Cell* **51**, 1079–1090.

Kalff-Suske M., Kunz J., Grzeschik K. H. and Suske G. (1995) Human Sp4 transcription factor gene (SP4) maps to chromosome 7p15. *Genomics* **26**, 631–633.

Kalff-Suske M., Kunz J., Grzeschik K-H. and Suske G. (1996) Human Sp3 transcriptional regulator gene (SP3) maps to chromosome 2q31. *Genomics* **37**, 410–412.

Karlseder J., Rotheneder H. and Wintersberger E. (1996) Interaction of Spl with the growth- and cell cycle-regulated transcription factor E2F. *Mol. Cell. Biol.* **16**, 1659–1667.

Kim S. J., Onwuta U. S., Lee Y. I., Li R., Botchan M. R. and Robbins P. D. (1992) The retinoblastoma gene product regulates Spl-mediated transcription. *Mol. Cell. Biol.* **12**, 2455–2463.

Kingsley C. and Winoto A. (1992) Cloning of GT box-binding proteins: a novel Spl multigene family regulating T-cell receptor gene expression. *Mol. Cell. Biol.* **12**, 4251–4261.

Latchman D. S. (1991) *Eukaryotic transcription factors.* Academic Press, London.

Licht J. D., Ro M., English M. A., Grossel M. and Hansen U. (1993) Selective repression of transcriptional activators at a distance by the Drosophila Kruppel protein. *Proc. Natl Acad. Sci. U.S.A.* **90**, 11361–11365.

Lin S. Y., Black A. R., Kostic D., Pajovic S., Hoover C. N. and Azizkhan J. C. (1996) Cell cycle-regulated association of E2F1 and Spl is related to their functional interaction. *Mol. Cell. Biol.* **16**, 1668–1675.

Macleod D., Charlton J., Mullins J. and Bird A. P. (1994) Spl sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes & Dev.* **8,** 2282–2292.

Marin M., Karis A., Visser P., Grosveld F. and Philipsen S. (1997) Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* **89,** 619–868.

Majello B., De Luca P., Hagen G., Suske G. and Lania L. (1994) Different members of the Sp1 multigene family exert opposite transcriptional regulation of the long terminal repeat of HIV-1. *Nucleic Acids Res.* **22,** 4914–4921.

Majello B., De Luca P. and Lania L. (1997) Sp3 is a bifunctional transcription regulator with modular independent activation and repression domains. *J. Biol. Chem.* **272,** 4021–4026.

Majello B., De Luca P., Suske G. and Lania L. (1995) Differential transcription regulation of c-myc promoter through the same DNA binding sites targeted by Sp1-like proteins. *Oncogene* **10,** 1841–1848.

Matera A. G. and Ward D. C. (1993) Localization of the human Sp1 transcription factor gene to 12q13 by fluorescence *in situ* hybridization. *Genomics* **17,** 793–794.

Mitchell P. J. and Tjian R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245,** 371–378.

Murata Y., Kim H. G., Rogers K. T., Udvadia A. J. and Horowitz J. M. (1994) Negative regulation of Sp1 trans-activation is correlated with the binding of cellular proteins to the amino terminus of the Sp1 trans-activation domain. *J. Biol. Chem.* **269,** 20674–20681.

Ohe N., Yamasaki Y., Sogawa K., Inazawa J., Ariyama T., Oshimura M. and Fujii-Kuriyama Y. (1993) Chromosomal localization and cDNA sequence of human BTEB, a GC box binding protein. *Somat. Cell Mol. Genet.* **19,** 499–503.

Pascal E. and Tjian R. (1991) Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes & Dev.* **5,** 1646–1656.

Perkins N. D., Edwards N. L., Duckett C. S., Agranoff A. B., Schmid R. M. and Nabel G. J. (1993) A cooperative interaction between NF-$_\kappa$B and Sp1 is required for HIV-1 activation. *EMBO J.* **12,** 3551–3558.

Prowse D. M., Bolgan L., Molnar A. and Dotto G. P. (1997) Involvement of the Sp3 transcription factor in induction of p21$^{cip1/WAF1}$ in keratinocyte differentiation. *J. Biol. Chem.* **272,** 1308–1314.

Saffer J. D., Jackson S. P. and Annarella M. B. (1991) Developmental expression of Sp1 in the mouse. *Mol. Cell. Biol.* **11,** 2189–2199.

Saffer J. D., Jackson S. P. and Thurston S. J. (1990) SV40 stimulates expression of the transacting factor Sp1 at the mRNA level. *Genes Dev.* **4,** 659–666.

Shao Z. and Robbins P. D. (1995) Differential regulation of E2F and Sp1-mediated transcription by G1 cyclins. *Oncogene* **10,** 221–228.

Sjottem E., Anderssen S. and Johansen T. (1996) The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3′ to the TATA box. *J. Virol.* **70,** 188–198.

Sogawa K., Imataka I., Yamasaki Y., Kusume H., Abe H. and Fujii-Kuriyama Y. (1993) cDNA cloning and transcriptional properties of a novel GC box-binding protein, BTEB2. *Nucleic Acids Res.* **21,** 1527–1532.

Supp D. M., Witte D. P., Branford W. W., Smith E. P. and Potter S. S. (1996) Sp4, a member of Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. *Dev. Biol.* **176,** 284–299.

Tate P., Skarnes W. and Bird A. (1996) The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nature Genet.* **12,** 205–208.

Udvadia A. J., Rogers K. T., Higgins P. D. R., Murata Y., Martin K. H., Humphrey P. A. and Horowitz J. M. (1993) Sp1 binds promoter elements regulated by RB protein and Sp1-mediated transcription is stimulated by RB coexpression. *Proc. Natl Acad. Sci. U.S.A.* **90,** 3265–3269.

Udvadia A. J., Templeton D. J. and Horowitz J. M. (1995) Functional interactions between the retinoblastoma (Rb) protein and Sp-family members: superactivation by Rb requires amino acids necessary for growth suppression. *Proc. Natl Acad. Sci. U.S.A.* **92,** 3953–3957.

Weinberg R. A. (1995) The retinoblastoma protein and cell cycle control. *Cell* **81,** 323–330.

Wimmer E. A., Jackle H., Pfeifle C. and Cohen S. M. (1993) A Drosophila homologue of human Sp1 is a head-specific segmentation gene. *Nature* **366,** 690–694.

Ye J., Xu R. H., Taylor-Papadimitriou J. and Pitha P. M. (1996) Sp1 binding plays a critical role in Erb-B2- and v-ras-mediated downregulation of α_2-integrin expression in human mammary epithelial cells. *Mol. Cell. Biol.* **16,** 6178–6189.

Zwicker J., Liu N., Engeland K., Lucibello F. C. and Muller R. (1996) Cell cycle regulation of E2F site occupation *in vivo*. *Science* **271,** 1595–1597.

D

ELSEVIER

Gene 238 (1999) 291–300

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Review

The Sp-family of transcription factors

G. Suske *

Institut für Molekularbiologie und Tumorforschung, Philipps-Universität Marburg, Emil-Mannkopff-Straße. 2, D-35037 Marburg, Germany

Received 14 June 1999; accepted 6 August 1999; Received by A.J. van Wijnen

---

Abstract

GC-boxes and related motifs are frequently occurring DNA-elements present in many promoters and enhancers. In contrast to other elements it was generally thought that the transcription factor Sp1 is the only factor acting through these motifs. The cloning of paralogous genes of the Sp1 factor uncovered the existence of a small protein family consisting of Sp1, Sp2, Sp3 and Sp4. All four proteins exhibit very similar structural features. They contain a highly conserved DNA-binding domain composed of three zinc fingers close the C-terminus and serine/threonine- and glutamine-rich domains in their N-terminal regions. The high degree of structural conservation between these four proteins suggested that they do exert similar functions. Molecular, genetic and biochemical analyses, however, demonstrated that Sp2, Sp3 and Sp4 are not simply functional equivalents of Sp1. Here, I will summarize and discuss recent advances which have been made towards understanding the mode of action and biological function of individual family members. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Review; Sp1; Sp2; Sp3; Sp4; Zinc finger

---

1. Introduction

Transcriptional regulation is exerted by the combinatorial action of proteins binding to distinct promoter and enhancer elements. Usually a limited number of *cis*-acting DNA elements is recognized not only by a single transcription factor but by a set of different proteins which are often structurally related (Latchman, 1995). Important and widely distributed promoter elements are G-rich elements such as the GC-box (GGGGCGGGG) and the related GT/CACCC-box (GGTGTGGGG). These elements are required for the appropriate expression of many ubiquitous, tissue-specific and viral genes. In addition, they occur frequently in the regulatory region of genes which are under a specific mode of control such as cell cycle regulation, hormonal activation and developmental patterning.

For some time it has been known that the general transcription factor Sp1 (Specificity protein 1) can bind to and act through the GC-boxes and it was generally accepted that this protein is an extremely versatile protein involved in the expression of many different genes documented by more than 2600 citations. More recently, however, it became clear that Sp1 is not the only protein acting through 'Sp1-binding sites' but simply represents the first identified and cloned protein of a small protein family. Currently this family consists of four proteins designated Sp1, Sp2, Sp3 and Sp4. Accordingly our view on Sp1 and its function has changed significantly. Here, I will summarize and discuss advances which have been directed towards understanding the properties and function of the individual Sp-proteins.

2. Molecular cloning of Sp transcription factors

Sp1 was originally identified as the transcription factor which binds to and activates transcription from multiple GC-boxes in the simian virus 40 (SV40) early promoter (Dynan and Tjian, 1983; Gidoni et al., 1984) and the thymidine kinase (TK) promoter (Jones et al., 1985). Molecular cloning of a partial human *Sp1* cDNA from HeLa cells was described in 1987 (Kadonaga et al., 1987). Since that time, the cDNA sequence published and deposited in databases is still incomplete. It encodes for the 696 C-terminal amino acids of human Sp1. The entire human Sp1 protein consists of 778 amino acids.

---

* Tel.: +49-6421-2866697. fax: +49-6421-2865398.
E-mail address: suske@imt.uni-marburg.de (G. Suske)

0378-1119/99/$ – see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0378-1119(99)00357-1

Table 1
Function and properties of Sp-factors

Factor	Accession numbers	Chromosomal localisation	Distribution	Transcriptional properties	Knockout phenotype	Features/ miscellaneous
Sp1	Human: J03133; Mouse: AF062566, AF022363; Rat: D12768	Human: 12q13 (Gaynor et al., 1993; Matera and Ward, 1993); Mouse: 15 (Saffer et al., 1990); Rat: 7q36 (Scohy et al., 1998)	Ubiquitous, developmental variations (Saffer et al., 1991)	Activator (Courey and Tjian, 1988) Synergistic activation (Courey et al., 1989) ; Superactivation (Pascal and Tjian, 1991)	Lethal at embryonic day 10 (Marin et al., 1997)	Two glutamine-rich activation domains (Courey and Tjian, 1988; Gill et al., 1994); phosphorylated (Jackson et al., 1990); glycosylated (Jackson and Tjian, 1988)
Sp2	Human: M97190, D28588	Human: 17q21.3-q22 (Scohy et al., 1998); Rat: 10q31-q32.1 (Scohy et al., 1998);	Various cell lines; tissues unknown (Kingsley and Winoto, 1992)	Unknown	Unknown	Original Sp2 clone incomplete; complete sequence in data bank D28588
Sp3	Human: X68560, S52144; Mouse: AF062567	Human: 2q31 (Kalff-Suske et al., 1996); Rat: 3q24-q31 (Scohy et al., 1998)	Ubiquitous (Hagen et al., 1992) and G. Suske, unpublished	Repressor of Sp1-mediated transcription (Hagen et al., 1994) Activator (Udvadia et al., 1995; Dennig et al., 1996)	Unknown	Two glutamine-rich activation domains; three isoforms (Hagen et al., 1994; Kennett et al., 1997); inhibitory domain (Dennig et al., 1996); translational start site of full length protein unknown
Sp4	Human: X68561, S50516; Mouse: U62522; Rat: U07610	Human: 7p15.3-p21 (Kalff-Suske et al., 1995); Mouse: 12 (Supp et al., 1996); Rat: 6q33 (Scohy et al., 1998)	Predominantly in neuronal cells; also in certain epithelia (Hagen et al., 1992; Supp et al., 1996) and G. Suske, unpublished	Activator (Hagen et al., 1994, 1995)	Growth retardation; males do not breed (Supp et al., 1996)	Two glutamine-rich activation domains (Hagen et al., 1995); entire human genomic sequence in data bank: Accession No. AC004595

Its sequence could be obtained from the Tjian laboratory upon request. More recently, highly conserved full length *Sp1* cDNAs from rat and mouse have been cloned (Imataka et al., 1992; Yajima et al., 1998) (Table 1).

The Sp1-related transcription factors Sp3 and Sp4 (originally also designated SPR-2 and SPR-1, respectively) were cloned by recognition site screening using the GT-box motif of the uteroglobin promoter as a probe (Hagen et al., 1992). Independently, Sp3 was obtained along with Sp2 from a human T-cell library by low stringency screening with the zinc finger-encoding region of human Sp1 as a probe (Kingsley and Winoto, 1992). The originally published human *Sp2* cDNA sequence assigned as complete coding sequence turned out to be incomplete. The complete coding sequence of *Sp2* is deposited in the data bank as human mRNA for the *KIAA0048* gene (Accession number D28588). The cDNAs coding for the murine homologues of Sp3 and Sp4 have been published recently (Supp et al., 1996; Yajima et al., 1998).

3. Structural features of Sp-family members

All four human Sp-family members have similar domain structures (Fig. 1A). They contain three zinc fingers close to the C-terminus and glutamine-rich domains adjacent to serine/threonine stretches in their N-terminal region. The 81 amino acids C2H2-type zinc finger region which represents the DNA-binding domain is the most highly conserved part of the proteins. Alignment of that region shows that Sp1, Sp3 and Sp4 are more closely related to each other than to Sp2 (Fig. 1B). According to structural studies on zinc finger proteins bound to DNA (Pavletich and Pabo, 1991; Fairall et al., 1993), one could predict that the amino acids KHA within the first, RER within the second and RHK within the third zinc finger contact specific bases. Especially these critical amino acids are all conserved in Sp1, Sp3 and Sp4 but not in Sp2. Consistently, Sp1, Sp3 and Sp4 recognize the classical Sp1-binding site with identical affinity (Hagen et al., 1992, 1994). In Sp2,

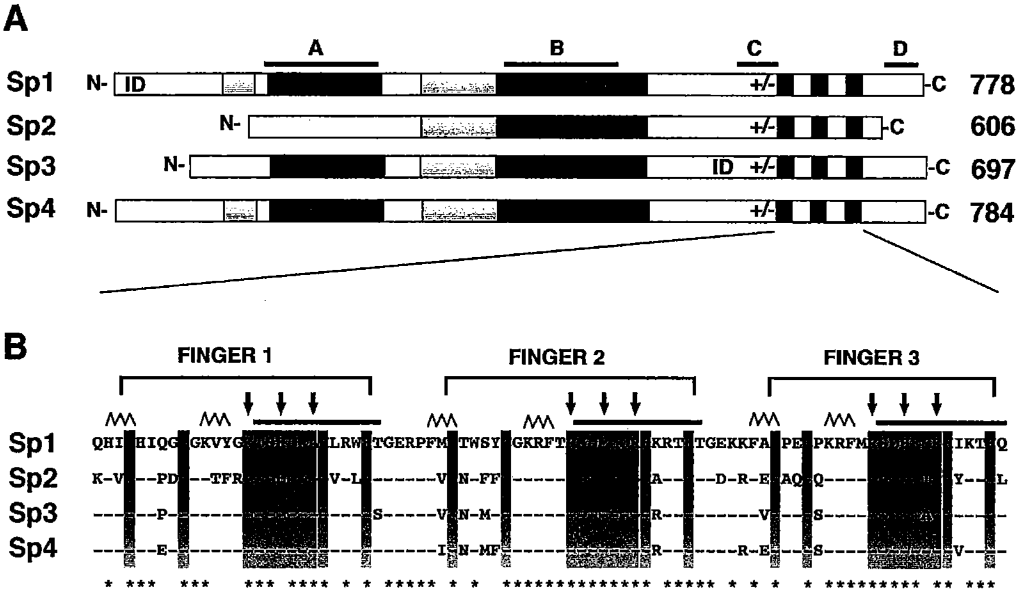

A
A
B
C
D
Sp1
N-
ID
+/-
-C
778
Sp2
N-
+/-
-C
606
Sp3
N-
ID
+/-
-C
697
Sp4
N-
+/-
-C
784
B
FINGER 1
FINGER 2
FINGER 3
Sp1
Sp2
Sp3
Sp4
QHI HIQG GKVYG LRW TGERPFM TWSY GKRFT KRT TGEKKFA PE PKRFM IKT Q
K-V --PD --TFR V-L -------V N-FF ----- A-- --D-R-E AQ Q---- Y-- L
-- --P- ------ ---S-----V N-M- ------ R-- ------V --S---- ----
-- --E- ------ ---------I N-MF ------ R-- ----R-E --S---- V--
* *** *** *** **** * * ***** * * ************** ***** * * * * ********* ** ***

Fig. 1. Structural features of Sp-proteins. (A) Schematic representation of the four human Sp-family members Sp1, Sp2, Sp3 and Sp4. Their length in amino acids is indicated on the right. The length of Sp3 refers to the full length isoform according to (Kingsley and Winoto, 1992). Colored boxes indicate regions of the proteins which are rich in glutamine (red) and serine/threonine (yellow) residues. The region preceding the first zinc finger (+/−) is rich in charged amino acids. The black boxes represent the zinc fingers. Lines above the draw of the Sp1 protein indicate the extent of four regions (A, B, C and D) which contribute to the transcriptional properties of Sp1 as defined by Tjian and coworkers (Courey and Tjian, 1988). Known activation (AD) and inhibitory domains (ID) are indicated. (B) Protein sequence alignment of the zinc finger domains. Stars below the sequence indicate sequence identity. Cysteine and histidine residues which coordinate zinc ions are underlayed in green and protein regions which contact the DNA in blue. Arrows point to the amino acids which determine the recognition specificity by contacting specific bases of the DNA (adapted from Pavletich and Pabo, 1991; Fairall et al., 1993). The black lines and the zig-zag lines indicate α-helical and β-sheet regions respectively.

the important histidine residue within the first zinc finger is replaced by a leucine residue (Fig. 1B). In accordance with this structural difference it was shown that Sp2 does not bind to the GC-box but to a GT-rich element within the T-cell receptor gene 5′-flanking region (Kingsley and Winoto, 1992).

4. Evolutionary relationship of the four Sp genes

The structural similarity of the four Sp-proteins suggests that they are evolutionally closely related. This is indeed the case and documented by their chromosomal localization in the human genome. All four *Sp* genes are found on paralogous chromosomal regions on human chromosomes 12q13 (*Sp1*), 17q21.3-q22 (*Sp2*), 2q31 (*Sp3*) and 7q21.3-q22 (*Sp4*) (Matera and Ward, 1993; Kalff-Suske et al., 1995, 1996; Scohy et al., 1998). The human *Sp* genes are linked to the homeobox gene cluster (O'Brien et al., 1993) on the corresponding chromosome (*Sp1/HOX C, Sp2/HOX B, Sp3/HOX D* and *Sp4/HOX A*). This chromosome configuration is maintained in rodents with the exception of *Sp4* which dissociated from the *Hox a* genes in the mouse and in the rat (Saffer et al., 1990; Supp et al., 1996; Scohy et al., 1998).

Except for the human *Sp4* gene, the sequence and the exon–intron structures of the individual chromosomal *Sp* genes are only partially known (Fig. 2). The entire sequence of the human *Sp4* gene is contained within the BAC clone RG023M10 (Accession No. AC004595). Cloning and mapping of the entire murine *Sp4* gene (G. Suske, unpublished data) revealed that the exon–intron structure of the *Sp4* gene is conserved between mouse and man. Partial gene structure information of the mouse *Sp1* (Chestier and Charnay, 1992; Marin et al., 1997) and *Sp3* genes (G. Suske, unpublished data) are also known (Fig. 2). The available data show that the exon–intron structures are conserved among individual *Sp*-family members supporting the conclusion that the four genes arise from a single ancestor gene.

Little is known about *Sp* genes in non-mammalian vertebrates and invertebrates. In vitro binding studies and biochemical characterizations have demonstrated the existence of GC-box binding proteins in fish which share immunological properties with Sp1 and Sp3 (Baudler et al., 1997). In *Drosophila melanogaster*, the head-specific protein buttonhead (btd) encodes a zinc finger type transcription factor with significant sequence and functional similarity to the Sp-proteins (Wimmer et al., 1993). However, the arrangement of the zinc finger region relative to the other domains is slighly different in btd. The region C-terminal to the zinc finger

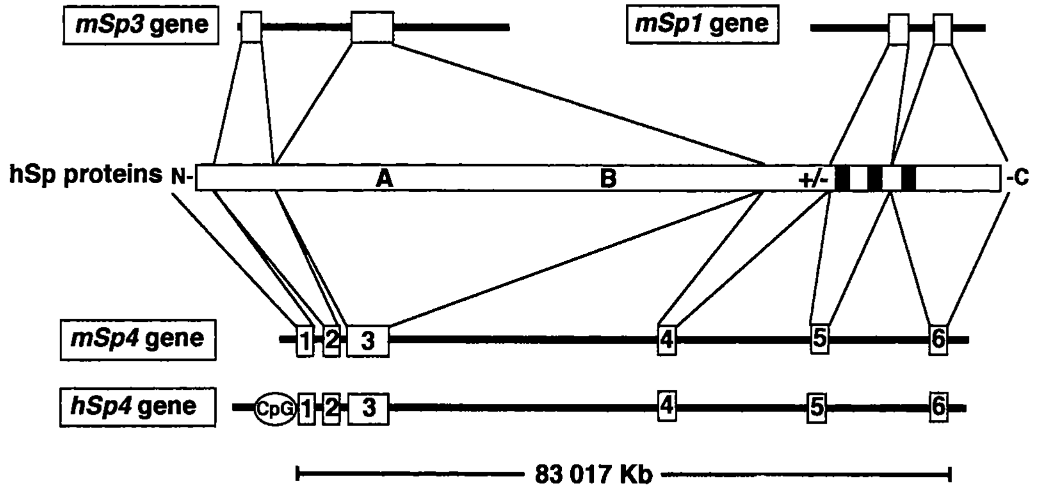


Fig. 2. Known genomic structures of *Sp1*, *Sp3* and *Sp4* genes substantiate their close evolutional relationship. The central drawing shows the structure of the Sp-proteins schematically. The glutamine-rich activation domains A and B and the zinc fingers are indicated. Known exons (numbered 1 to 6 in the case of the *Sp4* genes) of the mouse *Sp1*, *Sp3* and *Sp4* genes are connected with the corresponding protein regions. Both glutamine-rich activation domains are encoded on a single large exon. The human *Sp4* gene contains a CpG island in its 5′-flanking region.

region is extended and contains a serine/threonine-rich domain. Btd is required for development of the antennal, intercalary and mandibular segments of the fly head. A human *Sp1* transgene rescued the btd null phenotype partially. It supported development of the mandibular segment in the head of *btd* mutant embryos when expressed in the spatial pattern of btd (Wimmer et al., 1993). It would be interesting to know whether any of the other three Sp-proteins would have the capacity to rescue fully the btd null phenotype in flies. Since *Sp4* is expressed predominantly in the brain, it could be a good candidate gene.

5. Functional properties of individual Sp-family members

5.1. *Sp1: The prototype of the family*

Molecular cloning of Sp1 and its subsequent dissection revealed the functional domains of the protein. Both glutamine-rich regions (designated A and B) (Fig. 1A) can act as strong activation domains (Courey and Tjian, 1988). Mapping of the activation domain revealed that interspersed bulky hydrophobic amino acids are essential for transcriptional activation and not, per se, the glutamine residues (Gill et al., 1994). More recently, an inhibitory domain has been mapped to the N-terminus (Murata et al., 1994). Synergistic activation of promoters by Sp1 through multiple GC-boxes requires in addition the short C-terminal domain D (Pascal and Tjian, 1991). Sp1 is known to be phosphorylated (Jackson et al., 1990) and glycosylated (Jackson and Tjian, 1988), and it is capable of forming homotypic interactions leading to multimeric complexes (Mastrangelo et al., 1991; Pascal and Tjian, 1991; Su et al., 1991). In addition, many heterotypic interactions with different classes of nuclear proteins have been reported. These include factors belonging to the general transcription machinery, such as the TATA-box binding protein TBP (Emili et al., 1994) and the TBP-associated factors dTAFII110/hTAFII130 (Hoey et al., 1993; Tanese et al., 1996), and hTAFII55 (Chiang and Roeder, 1995). Other proteins which have been shown to interact with Sp1 are cell cycle regulators such as the retinoblastoma-related protein p107 (Datta et al., 1995) and transcription factors such as YY1 (Lee et al., 1993; Seto et al., 1993) or E2F (Karlseder et al., 1996; Lin et al., 1996). Sp1 can bind to its target sequence in assembled nucleosomes (Li et al., 1994), and in this regard it is interesting to note the interaction with a large coactivator complex called CRSP (cofactor required for Sp1 activation) which stimulates Sp1-mediated transcription in vitro (Ryu et al., 1999).

5.2. *Sp2*

There exists no functional analysis of Sp2 and correspondingly little is known about this Sp-family member. The presence of a potential glutamine-rich activation domain suggests that it may act as an activator. The T-cell antigen receptor α (*TCRα*) gene might be a target for Sp2 since it binds in vitro to a GT-box promoter element within the *TCRα* promoter (Kingsley and Winoto, 1992).

5.3. *Sp3: Activator versus repressor*

Unraveling the transcriptional role of Sp3 was complicated by the fact that three Sp3 isoforms exist, a 110-115 kDa Sp3 protein and two approximately

60-70 kDa Sp3 species. Very likely, the full length Sp3 protein is derived by translational initiation at a non-AUG which has been arbitrarily assigned to an AUU codon (Kingsley and Winoto, 1992). However, other potential non-AUG start codons are also present in that region and thus it remains to determine the N-terminus of Sp3. It has also been suggested that the published *Sp3* cDNA lacks 5′-untranslated sequences and the extreme N-terminus of the Sp3 protein. I consider this possibility unlikely since in vitro translation of the full length isoform of Sp3 from the published mRNA has been described (Kingsley and Winoto, 1992; Kennett et al., 1997). The two smaller Sp3 species arise from the first two internal AUG codons (G. Suske, unpublished observation, and Kennett et al., 1997). Consistently, antibodies raised against the N-terminus of Sp3 recognize in EMSAs the slow migrating complex which contains the full length Sp3 isoform but not the two fast migrating complexes which contain the two N-terminally truncated isoforms (Fig. 3).

Reports on the transcriptional properties of Sp3 appear, at first sight, contradictory. Sp3 has been shown to act as a transcriptional activator similar to Sp1 (Udvadia et al., 1995; Liang et al., 1996; Ihn and Trojanowska, 1997; Zhao and Chang, 1997, and many others). In other experiments, Sp3 remained inactive or acted only as a very weak activator (Hagen et al., 1994; Majello et al., 1994; Dennig et al., 1995; Kumar and Butler, 1997 and others). Most of these reports are based on co-transfection experiments into the insect cell line SL2. Usually, a distinct promoter fragment containing appropriate Sp-binding sites fused to a reporter gene was co-transfected along with Sp1 and Sp3 or both together. If Sp3 is expressed to the same extent as Sp1 but does not act as a strong activator, it will compete for the same binding site and thus lower Sp1-mediated activation. There are only a few reports in which additional approaches have been chosen to determine the potential role of Sp3 for the activity of a certain promoter (Noti, 1997; Hata et al., 1998). An antisense strategy which knocked out endogenous Sp3 expression in the myelomonocytic cell line HL60 revealed that Sp3 participates in the activation of the *CD11c* and *CD11b* promoters (Noti, 1997).

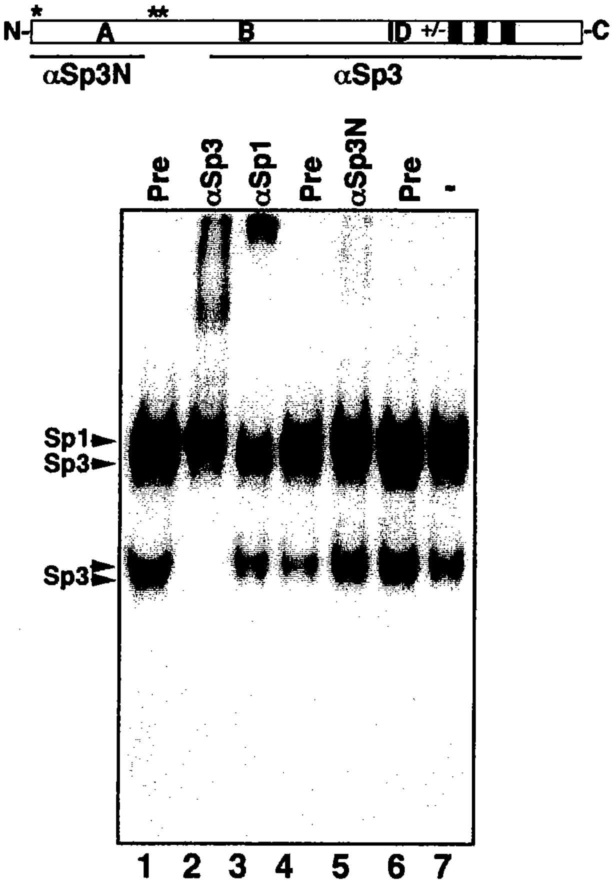


Fig. 3. Sp1 and the three isoforms of Sp3 are visible in an electrophoretic mobility shift assay (EMSA). A schematic representation of Sp3 is shown at the top. The activation domains A and B, the inhibitory domain (ID), the charged region (+/−) preceding the DNA-binding domain and the three zinc fingers (black boxes) are depicted. Asterisks indicate the N-termini of the three isoforms. The black lines indicate the segments of Sp3 which have been used for generation of the antibodies αSp3N and αSp3. The EMSA was performed with a ^{32}P-labeled GC-box oligonucleotide and nuclear extract from CV-1 cells. Antisera against Sp1 (αSp1, lane 3), the N-terminal part of Sp3 (αSp3N, lane 5), the C-terminal part of Sp3 (αSp3, lane 2) or the corresponding preimmune sera (Pre, lanes 1, 4 and 6), respectively, were included in the binding reactions as indicated. Complexes containing Sp1 or Sp3 are indicated at the left. The antiserum directed to the C-terminus shifted all three Sp3-containing complexes (lane 2) whereas the antiserum directed to the N-terminus shifted only the slow migrating complex (lane 5).

The experimental conditions which are needed for Sp3 to act as a strong activator or a transcriptional inactive molecule which represses Sp1-mediated activation are not completely understood. The structure and the arrangement of the recognition sites appear to determine whether Sp3 is transcriptionally inactive and can repress Sp1-mediated activation or whether it acts as a strong activator. Promoters containing a single binding site are activated, whereas promoters containing multiple binding sites often are not activated or respond weakly to Sp3 (Birnbaum et al., 1995; Dennig et al., 1996). Whether Sp3 acts as an activator or as a repressor of Sp1-mediated activation might also depend of the cellular context. Transfected Sp3 stimulated transcription from the HERV-H long-terminal repeat in the teratocarcinoma cell line NTera2-D1 but acted as a repressor in HeLa and insect cells (Sjottem et al., 1996).

It has been suggested that the two small Sp3 isoforms might act as repressor molecules whereas the full length Sp3 isoform does act as an activator (Kennett et al., 1997). Although attractive and simple, this model does not seem to hold true. Exclusive expression of full length Sp3 triggered by an artificial leader sequence can also repress Sp1-mediated activation (Dennig et al., 1996).

It is clear that both N-terminal glutamine-rich regions can act as strong activation domains on their own in both insect and in mammalian cells (Dennig et al., 1996; Majello et al., 1997). The molecular basis for the inactivity of Sp3 under certain conditions has been mapped to an inhibitory domain located between the second glutamine-rich activation domain and the first zinc finger. The amino acid triplet KEE within this domain is absolutely essential for repressor function (Dennig et al., 1996). Mutation of these amino acids to alanine residues converted almost inactive Sp3 to a strong activator. The inhibitory domain of Sp3 acts as an independent modul in *cis*. It can be transferred to other activation domains which in turn lose their activation properties (Dennig et al., 1996). Yet, we do not know how this domain functions mechanistically. Purified recombinant Sp3 expressed in SL2 cells (Braun and Suske, 1999) act in an in vitro system as strong activator similar to Sp1 (H. Braun and G. Suske, unpublished). From this, and other observations, it can be concluded that additional proteins which act as co-repressors are involved in the inhibitory function of this domain. Our recent cloning of a protein designated SIF-1 (Sp3-interacting protein 1) which specifically interacted with the wild type inhibitory domain but not with the mutated form support this idea (A. Doll and G. Suske, unpublished).

5.4. Sp4: The tissue-specific Sp-family member

Sp4 was cloned along with Sp3 by virtue of its binding to the GT-box of the uteroglobin promoter (Hagen et al., 1992). In contrast to Sp1 and Sp3 which are ubiquitous transcription factors, Sp4 expression appears to be restricted to a few tissues. High levels of Sp4 are predominantly found in the brain (Hagen et al., 1992; Supp et al., 1996). Consistent with its high expression in the central nervous system, targeted disruption of the mouse Sp4 gene by homologous recombination led to behavioral defects (see Section 7).

Sp4 exhibits specific functional properties distinct from Sp1 and Sp3. The transactivation function of Sp4 resides, like that of Sp1, in the N-terminal glutamine-rich regions. In contrast to Sp1, Sp4 is not able to act synergistically through adjacent binding sites although Sp4 can function as a target for the Sp1 activation domains in a superactivation assay suggesting that the activation domains of Sp1 and Sp4 are functionally related (Hagen et al., 1995).

6. Regulation by the ratio of Sp1 and Sp3

It is clear that in a given cell type, co-expression of Sp1 and Sp3 occurs and we have to assume that these two proteins compete for the same binding sites in vivo. The initial characterization of Sp1 and Sp3 shows that they differ in their capacity to activate or repress transcription. Independently of whether Sp3 acts as an activator or as a repressor of Sp1-mediated activation, the relative abundance of Sp1 and Sp3 should allow regulation of gene activities. The abundance of Sp1 and Sp3 varies among different cell types. In endothelial cells which contain high levels of both Sp1 and Sp3, the Sp1/Sp3 ratio is higher than in non-endothelial cells (Hata et al., 1998). The endothelial-specific activity of the *KDR/flk-1* promoter was mapped to an Sp1/Sp3-binding site, and Sp3 attenuated Sp1-mediated *KDR/flk-1* promoter activation. Thus, it was suggested that endothelial selective *KDR/flk-1* expression may be partially mediated by the high Sp1/Sp3 ratio in these cells (Hata et al., 1998).

Alterations in the relative abundance of Sp1 and Sp3 upon different conditions have been reported in several cases. In primary keratinocytes, Sp3 levels exceed those of Sp1. The Sp3/Sp1 ratio beomes inverted if these cells differentiate in vitro on treatment with high calcium suggesting that cell type differential transcription of several genes might be regulated by Sp1 and Sp3 (Apt et al., 1996). Worth mentioning in this context is that in differentiated keratinocytes only Sp3, but not Sp1, overexpression enhanced the induction of the *p21* promoter (Prowse et al., 1997).

A change in the Sp1/Sp3 ratio occurred also when C2C12 myocytes were cultivated under hypoxic conditions. Hypoxia caused a progressive depletion of Sp3 whereas the Sp1 protein level remained unchanged (Discher et al., 1998). Like other glycolytic enzymes the muscle-specific pyruvate kinase-M and β-enolase are upregulated under low oxygen pressure. The hypoxia responsive promoter elements in these two genes have been mapped to GC-rich elements bound by Sp1 and Sp3. Thus, it was concluded that hypoxia activates these glycolytic enzyme gene promoters by down-regulating Sp3, thereby removing the associated transcriptional repression (Discher et al., 1998).

In another report, it was shown that downregulation of *α2-integrin* gene expression in mammary epithelial cells by Erb-B2 and v-Hras is mediated by two adjacent Sp1/Sp3 binding sites in the *α2-integrin* promoter. In that case, however, the Sp3 DNA-binding activity remained unaltered whereas the Sp1 DNA-binding activity was reduced (Ye et al., 1996).

Although not shown so far, one could imagine that alterations of the transcriptional capacity of Sp1 or Sp3 could be obtained by other mechanisms. Especially, Sp3 could be a target of signal transduction pathways since it has both activation and repression functions. Modifications in the protein might alter the potency of the protein in one or the other direction. Altogether, further unraveling of the molecular structure and modifications of the protein should give insights into the specific role of Sp3.

7. Physiological function of Sp-proteins

In the past, a large variety of biological functions have been assigned to Sp1-binding sites and to Sp1. However, the identification of the three paralogous proteins Sp2, Sp3 and Sp4 raises the question as to which tasks are performed by which protein. This question is particularly interesting for Sp1 and Sp3 because both proteins are present in the same cell and are indistinguishable in their DNA-binding specificity. Gene disruption in mice is a powerful tool for obtaining information on specific functions of individual Sp-proteins.

Given that Sp1 is implicated in the activation of a very large number of genes, such as housekeeping, tissue-specific and cell cycle-regulated genes, and is required to prevent methylation of CpG islands (Brandeis et al., 1994; Macleod et al., 1994), one would expect that cells lacking Sp1 would not survive. Surprisingly, this is not the case. S. Philipsen and his coworkers have disrupted the mouse *Sp1* gene and found that Sp1-deficient embryonic stem cells (ES cells) are viable, have normal growth characteristics and can be induced to differentiate and form embryoid bodies as efficiently as wild type ES cells (Marin et al., 1997). Nevertheless, Sp1 is essential for normal mouse embryogenesis. The *Sp1*-knockout embryos are severely retarded in development and they all died around day 11 of gestation. They displayed a marked heterogeneity in phenotype indicating that Sp1 has indeed a general function in many cell types. Interestingly, the defects in *Sp1-/-* mice are caused by a cell autonomous mechanism. *Sp1-/-* ES cells injected into blastocysts contributed efficiently to chimaeric embryos at early stages but after day E 11 they rapidly declined with no contribution to newborn mice. Thus, Sp1 appears to be a transcription factor whose function is essential for differentiated cells after day10 of development.

Independently of the severe developmental defects of the Sp1 null mice, the embryos express many putative Sp1 target genes at normal levels, including housekeeping and cell-cycle regulated genes. In addition, CpG-islands remained methylation free. So far, the only genes which were found to be expressed at a lower level in *Sp1-/-* mice are the thymidine kinase and the methyl-CpG binding protein 2 (*MeCP2*) genes (Marin et al., 1997). Since there are striking similarities between the phenotypes of the *MeCP2* (Tate et al., 1996) and the *Sp1* knockouts, it was suggested that the *MeCP2* gene might be a key target of Sp1. However, whether Sp1 acts as a direct regulator of MeCP2 expression by binding to promoter, enhancer or local control region elements in the *MeCP2* gene, or whether additional proteins mediate downregulation of MeCP2 remains to be established.

The *Sp1*-knockout demonstrated that Sp1 is not essential for the expression of many genes previously shown be activated in cell culture transfection experiments. One could speculate that other Sp-family members compensate at early embryonic stages, at least in part for the loss of Sp1 activity. Sp3 would be a good candidate because it is also ubiquitously expressed, has the potential to activate transcription and its DNA-binding activity is indistinguishable from Sp1. In that regard, the phenotype of *Sp3* knockout mice will be of great interest.

Sp4, the tissue restricted member of the Sp-family, is predominantly expressed in the brain but also detectable in epithelial tissues, testis and developing teeth (Hagen et al., 1992; Supp et al., 1996). Disruption of the mouse *Sp4* gene revealed that it is important for early post-natal survival (Supp et al., 1996). Approximately two thirds of the *Sp4-/-* mice die within a few days of birth. The cause of the early death remains unknown. Those mice which survive are significantly smaller than their wild type littermates. The reduced body weight appears to result from an unknown, but growth hormone independent, mechanism (Supp et al., 1996). Interestingly, surviving mice exhibit a striking sex-specific abnormality. While fertility of the female mutants appears normal, males do not breed although their reproductive organs are fully developed and apparently normal. It appears that male *Sp4-/-* mice are unable to copulate. The most likely cause of this abnormal behavior is a neurological defect. The hypothalamus and the vomeronasal organ are known to play important roles in reproductive physiology and behavior. However, both structures are histologically normal in *Sp4-/-* mice. Thus, we await further investigation to understand the role of Sp4 and to identify its target genes.

In that context, it should be noted that for both the *Sp1* and the *Sp4* knockouts, the zinc finger regions have been chosen to inactivate the genes. In both cases the N-terminal part encoding the transactivation domains are still expressed (Supp et al., 1996; Marin et al., 1997). One could speculate that the activation domains on their own might act as a gain-of-function or might interfere with other Sp-family members. In the case of the *Sp4* knockout such a scenario does not seem to be the case. Knockout *Sp4-/-* mice which do not express the N-terminal part of the protein manifest the same phenotype (G. Suske, unpublished observations).

8. Yet more and more GC/GT-box binding proteins

In the past, essential GC-boxes in promoters were often equated with “Sp1-binding sites’ thereby overlooking the fact that Sp1 is not the only protein which recognizes this important element. In addition to Sp3 and Sp4 there exist at least three other proteins, BTEB1 (basic transcription element binding protein 1) (Imataka et al., 1992), TIEG1 and TIEG2 (TGFβ-inducible early protein genes 1 and 2) (Subramaniam et al., 1995; Cook et al., 1998; Fautsch et al., 1998) which have a binding specificity very similar, if not identical, to Sp1 (Sogawa et al., 1993b; Cook et al., 1998). In BTEB1, TIEG1 and TIEG2 the DNA-binding domain is found also close to the C-terminus and consists of three C2H2-type zinc fingers. Most significantly, all amino acids which are believed to specifically contact the DNA are conserved between these three proteins and Sp1, Sp3 and Sp4. However, the N-terminal domains of these proteins do not share any similarity with the Sp-proteins. BTEB1 contains acidic and hydrophobic amino acid stretches, and TIEG1 as well as the closely related protein TIEG2 are rich in proline residues.

In addition, another subgroup of zinc finger proteins, including the so called Krüppel-like factors, has to be considered when essential 'Sp1-binding sites' are analyzed in promoters. These proteins also contain three zinc fingers close to their C-terminus and bind preferentially to so called GT-or CACCC-boxes. Sp-family members also recognize these elements. Thus, Sp-proteins and Krüppel-like factors have an overlapping binding specificity. Currently this subfamily of zinc finger proteins consists of at least 10 different members, the so called Krüppel like factors EKLF (Miller and Bieker, 1993), BKLF/TEF-2 (Crossley et al., 1994), GKLF (Shields et al., 1996), IKLF (Conkright et al., 1999), LKLF (Anderson et al., 1995), UKLF (Matsumoto et al., 1998) and FKLF (Asano et al., 1999) and ZNF741 (only sequence in data base), AP-2rep (Imhof et al., 1999) and ZF9 (Ratziu et al., 1998). Another factor called BTEB2 (Sogawa et al., 1993a) is very likely an N-terminally truncated human ortholog of murine IKLF. Understanding the biological function and the mode of action of all these GC/GT-box binding proteins will be a scientific challenge.

9. Conclusions

The cloning and initial characterization of a family of Sp-proteins as well as other GC/GT-box binding proteins provides much information about the potential functions and activities of these proteins. The most obvious question, however, concerns the specificity of the individual family members especially of Sp1 and Sp3. Neither binding site preferences nor differential expression patterns seem to confer specificity of these two proteins. However, activation of a given promoter requires the binding of multiple transcription factors which might bind cooperatively to their sites or might act synergistically by other mechanisms. Thus far, little is known how Sp1 and Sp3 act on natural promoters in combination with other transcription factors in vivo. Specificity could be obtained also by the interaction with co-activators or co-repressors. In that regard it might be interesting to know whether the co-activator complex CRSP (Ryu et al., 1999) is specific for Sp1 or whether it can also cooperate with other Sp-family members.

The most interesting question at this stage concerns the key target genes for Sp1 and Sp3. The Sp1-knockout suggests that there are only a few genes which are regulated by Sp1. However, only a few genes have actually been tested. New techniques such as DNA microchip arrays could be helpful in identifying additional genes whose expression is dependent on Sp1. On the other hand, knocking out a single family member may not reflect the whole truth because overlapping functions with other Sp-family members might conceal important in vivo functions. In addition, the *Sp1*-knockout mice do not provide information on Sp1 in differentiated tissues in the adult animal. For that we have to await for the conditional disruption of the *Sp1* gene in specific tissues and at specific stages during development.

Acknowledgements

I thank H. Braun and H. Göllner for discussions and M. Kalff-Suske for critically reading the manuscript. The Deutsche Forschungsgemeinschaft and the Stiftung P.E. Kempkes financially support research in my group.

References

Anderson, K.P., Kern, C.B., Crable, S.C., Lingrel, J.B., 1995. Isolation of a gene encoding a functional zinc finger protein homologous to erythroid Krüppel-like factor: identification of a new multigene family. Mol. Cell. Biol. 15, 5957–5965.

Apt, D., Watts, R.M., Suske, G., Bernard, H.U., 1996. High Sp1/Sp3 ratios in epithelial cells during epithelial differentiation and cellular transformation correlate with the activation of the HPV-16 promoter. Virology 224, 281–291.

Asano, H., Li, X.S., Stamatoyannopoulos, G., 1999. FKLF, a novel Krüppel-like factor that activates human embryonic and fetal beta-like globin genes. Mol. Cell. Biol. 19, 3571–3579.

Baudler, M., Duschl, J., Winkler, C., Schartl, M., Altschmied, J., 1997. Activation of transcription of the melanoma inducing Xmrk oncogene by a GC box element. J. Biol. Chem. 272, 131–137.

Birnbaum, M.J., van Wijnen, A.J., Odgren, P.R., Last, T.J., Suske, G., Stein, G.S., Stein, J.L., 1995. Sp1 trans-activation of cell cycle regulated promoters is selectively repressed by Sp3. Biochemistry 34, 16 503–16 508.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., Cedar, H., 1994. Sp1 elements protect a CpG island from de novo methylation. Nature 371, 435–438.

Braun, H., Suske, G., 1999. Vectors for inducible expression of dual epitope-tagged proteins in insect cells. BioTechniques. in press

Chestier, A., Charnay, P., 1992. Difference in the genomic organizations of the related transcription factors Sp1 and Krox-20, possible evolutionary significance. DNA Sequence 2, 325–327.

Chiang, C.M., Roeder, R.G., 1995. Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. Science 267, 531–536.

Conkright, M.D., Wani, M.A., Anderson, K.P., Lingrel, J.B., 1999. A gene encoding an intestinal-enriched member of the Krüppel-like factor family expressed in intestinal epithelial cells. Nucleic Acids Res. 27, 1263–1270.

Cook, T., Gebelein, B., Mesa, K., Mladek, A., Urrutia, R., 1998. Molecular cloning and characterization of TIEG2 reveals a new subfamily of transforming growth factor-ß-inducible Sp1-like zinc finger-encoding genes involved in the regulation of cell growth. J. Biol. Chem. 273, 25 929–25 936.

Courey, A.J., Holtzman, D.A., Jackson, S.P., Tjian, R., 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell 59, 827–836.

Courey, A.J., Tjian, R., 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell 55, 887–898.

Crossley, M., Tsang, A.P., Bieker, J.J., Orkin, S.H., 1994. Regulation of the erythroid Krüppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. J. Biol. Chem. 269, 15 440–15 444.

Datta, P.K., Raychaudhuri, P., Bagchi, S., 1995. Association of p107 with Sp1: genetically separable regions of p107 are involved in regulation of E2F- and Sp1-dependent transcription. Mol. Cell. Biol. 15, 5444–5452.

Dennig, J., Beato, M., Suske, G., 1996. An inhibitor domain in Sp3 regulates its glutamine-rich activation domains. EMBO J. 15, 5659–5667.

Dennig, J., Hagen, G., Beato, M., Suske, G., 1995. Members of the Sp transcription factor family control transcription from the uteroglobin promoter. J. Biol. Chem. 270, 12 737–12 744.

Discher, D.J., Bishopric, N.H., Wu, X., Peterson, C.A., Webster, K.A., 1998. Hypoxia regulates beta-enolase and pyruvate kinase-M promoters by modulating Sp1/Sp3 binding to a conserved GC element. J. Biol. Chem. 273, 26 087–26 093.

Dynan, W.S., Tjian, R., 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35, 79–87.

Emili, A., Greenblatt, J., Ingles, C.J., 1994. Species-specific interaction of the glutamine-rich activation domains od Sp1 with the TATA box-binding protein. Mol. Cell. Biol. 14, 1582–1593.

Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T., Rhodes, D., 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature 366, 483–487.

Fautsch, M.P., Vrabel, A., Subramaniam, M., Hefferen, T.E., Spelsberg, T.C., Wieben, E.D., 1998. TGFβ-inducible early gene (TIEG) also codes for early growth response alpha (EGRα): evidence of multiple transcripts from alternate promoters. Genomics 51, 408–416.

Gaynor, R.B., Shieh, B.H., Klisak, I., Sparkes, R.S., Lusis, A.J., 1993. Localization of the transcription factor SP1 gene to human chromosome 12q12-->q13.2. Cytogenet. Cell. Genet. 64, 210–212.

Gidoni, D., Dynan, W.S., Tjian, R., 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature 312, 409–413.

Gill, G., Pascal, E., Tseng, Z.H., Tjian, R, 1994. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAF$_{II}$110 component of the Drosophila TFIID complex and mediates transcriptional activation. Proc. Natl. Acad. Sci. USA 91, 192–196.

Hagen, G., Dennig, J., Preiss, A., Beato, M., Suske, G., 1995. Functional analyses of the transcription factor Sp4 reveal properties distinct from Sp1 and Sp3. J. Biol. Chem. 270, 24 989–24 994.

Hagen, G., Müller, S., Beato, M., Suske, G., 1992. Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. Nucleic Acids Res. 20, 5519–5525.

Hagen, G., Müller, S., Beato, M., Suske, G., 1994. Sp1-mediated transcriptional activation is repressed by Sp3.. EMBO J. 13, 3843–3851.

Hata, Y., Duh, E., Zhang, K., Robinson, G.S., Aiello, L.P., 1998. Transcription factors Sp1 and Sp3 alter vascular endothelial growth factor receptor expression through a novel recognition sequence. J. Biol. Chem. 273, 19 294–19 303.

Hoey, T., Weinzierl, R.O.J., Gill, G., Chen, J.-L., Dynlacht, B.D., Tjian, R., 1993. Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators. Cell 72, 247–260.

Ihn, H., Trojanowska, M., 1997. Sp3 is a transcriptional activator of the human alpha2(I) collagen gene.. Nucleic Acids Res. 25, 3712–3717.

Imataka, H., Sogawa, K., Yasumoto, K., Kikuchi, Y., Sasano, K., Kobayashi, A., Hayami, M., Fujii-Kuriyama, Y., 1992. Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene. EMBO J. 11, 3663–3671.

Imhof, A., Schuierer, M., Werner, O., Moser, M., Roth, C., Bauer, R., Buettner, R., 1999. Transcriptional regulation of the AP-2α promoter by BTEB-1 and AP- 2rep, a novel wt-1/egr-related zinc finger repressor. Mol. Cell. Biol. 19, 194–204.

Jackson, S.P., MacDonald, J.J., Lees-Miller, S., Tjian, R., 1990. GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. Cell 63, 155–165.

Jackson, S.P., Tjian, R., 1988. O-Glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. Cell 55, 125–133.

Jones, K.A., Yamamoto, K.R., Tjian, R., 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell 42, 559–572.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R., Tjian, R., 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51, 1079–1090.

Kalff-Suske, M., Kunz, J., Grzeschik, K.-H., Suske, G., 1995. Human Sp4 transcription factor gene (SP4) maps to chromosome 7p15. Genomics 26, 631–633.

Kalff-Suske, M., Kunz, J., Grzeschik, K.-H., Suske, G., 1996. Human Sp3 transcriptional regulator gene (SP3) maps to chromosome 2q31. Genomics 37, 410–412.

Karlseder, J., Rotheneder, H., Wintersberger, E., 1996. Interaction of Sp1 with the growth- and cell cycle-regulated transcription factor E2F. Mol. Cell. Biol. 16, 1659–1667.

Kennett, S.B., Udvadia, A.J., Horowitz, J.M., 1997. Sp3 encodes multiple proteins that differ in their capacity to stimulate or repress transcription. Nucleic Acids Res. 25, 3110–3117.

Kingsley, C., Winoto, A., 1992. Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. Mol. Cell. Biol. 12, 4251–4261.

Kumar, A.P., Butler, A.P., 1997. Transcription factor Sp3 antagonizes activation of the ornithine decarboxylase promoter by Sp1. Nucleic Acids Res. 25, 2012–2019.

Latchman, D.S., 1995. in: Eukaryotic Transcription Factors, 2nd Ed, Academic Press, London.

Lee, J.S., Galvin, K.M., Shi, Y., 1993. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. Proc. Natl. Acad. Sci. USA 90, 6145–6149.

Li, B., Adams, C.C., Workman, J.L., 1994. Nucleosome binding by the constitutive transcription factor Sp1. J. Biol. Chem. 269, 7756–7763.

Liang, Y., Robinson, D.F., Dennig, J., Suske, G., Fahl, W.E., 1996. Transcriptional regulation of the SIS/PDGF-B gene in human osteosarcoma cells by the Sp family of transcription factors. J. Biol. Chem. 271, 11 792–11 797.

Lin, S.-Y., Black, A.R., Kostic, D., Pajovic, S., Hoover, C.N., Azizkhan, J.C., 1996. Cell cycle-regulated association of E2F and Sp1 is related to their functional interaction. Mol. Cell. Biol. 16, 1668–1675.

Macleod, D., Charlton, J., Mullins, J., Bird, A.P., 1994. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. 8, 2282–2292.

Majello, B., De Luca, P., Hagen, G., Suske, G., Lania, L., 1994. Different members of the Sp1 multigene family exert opposite transcriptional regulation of the long terminal repeat of HIV-1. Nucleic Acids Res. 22, 4914–4921.

Majello, B., De Luca, P., Lania, L., 1997. Sp3 is a bifunctional transcription regulator with modular independent activation and repression domains. J. Biol. Chem. 272, 4021–4026.

Marin, M., Karis, A., Visser, P., Grosveld, F., Philipsen, S., 1997. Transcription factor Sp1 is essential for early development but dispensable for cell growth and differentiation. Cell 89, 619–628.

Mastrangelo, I.A., Courey, A.J., Wall, J.S., Jacsdon, S.P., Hough, P.V.C, 1991. DNA looping and Sp1 multimer links: A mechanism for transcriptional synergism and enhancment. Proc. Natl. Acad. Sci. USA 88, 5670–5674.

Matera, A.G., Ward, D.C., 1993. Localization of the human Sp1 transcription factor gene to 12q13 by fluorescence in situ hybridization. Genomics 17, 793–794.

Matsumoto, N., Laub, F., Aldabe, R., Zhang, W., Ramirez, F., Yoshida, T., Terada, M., 1998. Cloning the cDNA for a new human zinc finger protein defines a group of closely related Krüppel-like transcription factors. J. Biol. Chem. 273, 28 229–28 237.

Miller, I.J., Bieker, J.J., 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. Mol. Cell. Biol. 13, 2776–2786.

Murata, Y.H.G.K., Rogers, K.T., Udvadia, A.J., Horowitz, J.M., 1994. Negative regulation of Sp1 trans-activation is correlated with the binding of cellular proteins to the amino terminus of the Sp1 trans-activation domain. J. Biol. Chem. 269, 20 674–20 681.

Noti, J.D., 1997. Sp3 mediates transcriptional activation of the leukocyte integrin genes CD11C and CD11B and cooperates with c-Jun to activate CD11C. J. Biol. Chem. 272, 24 038–24 045.

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A., Copeland, N.G., 1993. Anchored reference loci for comparative genome mapping in mammals. Nat. Genet. 3, 103–112.

Pascal, E., Tjian, R., 1991. Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. Genes Dev. 5, 1646–1656.

Pavletich, N.P., Pabo, C.O., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. Science 252, 809–817.

Prowse, D.M., Bolgan, L., Molnar, A., Dotto, G.P., 1997. Involvement of the Sp3 transcription factor in induction of p21Cip1/WAF1 in keratinocyte differentiation. J. Biol. Chem. 272, 1308–1314.

Ratziu, V., Lalazar, A., Wong, L., Dang, Q., Collins, C., Shaulian, E., Jensen, S., Friedman, S.L., 1998. Zf9, a Krüppel-like transcription factor up-regulated in vivo during early hepatic fibrosis. Proc. Natl. Acad. Sci. USA 95, 9500–9505.

Ryu, S., Zhou, S., Ladurner, A.G., Tjian, R., 1999. The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. Nature 397, 446–450.

Saffer, J.D., Jackson, S.P., Annarella, M.B., 1991. Developmental expression of Sp1 in the mouse. Mol. Cell Biol. 11, 2189–2199.

Saffer, J.D., Thurston, S.J., Annarella, M.B., Compton, J.G., 1990. Localization of the gene for the trans-acting transcription factor SP1 to the distal end of mouse chromosome 15. Genomics 8, 571–574.

Scohy, S., Van Vooren, P., Szpirer, C., Szpirer, J., 1998. Assignment of Sp genes to rat chromosome bands 7q36 (Sp1), 10q31-- >q32.1 (Sp2), 3q24-- >q31 (Sp3) and 6q33 (Sp4) and of the SP2 gene to human chromosome bands 17q21.3-- >q22 by in situ hybridization. Cytogenet. Cell. Genet. 81, 273–274.

Seto, E., Lewis, B., Shenk, T., 1993. Interaction between transcription factors Sp1 and YY1. Nature 365, 462–464.

Shields, J.M., Christy, R.J., Yang, V.W., 1996. Identification and characterization of a gene encoding a gut-enriched Krüppel-like factor expressed during growth arrest. J. Biol. Chem. 271, 20 009–20 017.

Sjottem, E., Anderssen, S., Johansen, T., 1996. The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3′ to the TATA box. J. Virol. 70, 188–198.

Sogawa, K., Imataka, H., Yamasaki, Y., Kusume, H., Abe, H., Fujii-Kuriyama, Y., 1993a. cDNA cloning and transcriptional properties of a novel GC box-binding protein, BTEB2. Nucleic Acids Res. 21, 1527–1532.

Sogawa, K., Kikuchi, Y., Imataka, H., Fujii-Kuriyama, Y., 1993b. Comparison of DNA-binding properties between BTEB and Sp1. J. Biochem. Tokyo 114, 605–609.

Su, W., Jackson, S., Tjian, R., Echols, H., 1991. DNA looping between sites for transcriptional activation: self-association of DNA-bound SP1. Genes Dev. 5, 820–826.

Subramaniam, M., Harris, S.A., Oursler, M.J., Rasmussen, K., Riggs, B.L., Spelsberg, T.C., 1995. Identification of a novel TGFβ-regulated gene encoding a putative zinc finger protein in human osteoblasts. Nucleic Acids Res. 23, 4907–4912.

Supp, D.M., Witte, D.P., Branford, W.W., Smith, E.P., Potter, S.S., 1996. Sp4, a member of the Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. Dev. Biol. 176, 284–299.

Tanese, N., Saluja, D., Vassallo, M.F., Chen, J.L., Admon, A., 1996. Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100. Proc. Natl. Acad. Sci. USA 93, 13 611–13 616.

Tate, P., Skarnes, W., Bird, A., 1996. The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. Nat. Genet. 12, 205–208.

Udvadia, A.J., Templeton, D.J., Horowitz, J.M., 1995. Functional interactions between the retinoblastoma (Rb) protein and Sp-family members: superactivation by Rb requires amino acids necessary for growth suppression. Proc. Natl. Acad. Sci. USA 92, 3953–3957.

Wimmer, E.A.Jäckle, Wimmer, E.A., Jäckle, H., Pfeifle, C., Cohen, S.M., 1993. A *Drosophila* homologue of human Sp1 is a head-specific segmentation gene. Nature 366, 690–694.

Yajima, S., Lee, S.H., Minowa, T., Mouradian, M.M., 1998. Sp family transcription factors regulate expression of rat D2 dopamine receptor gene. DNA Cell. Biol. 17, 471–479.

Ye, J., Xu, R.H., Taylor-Papadimitriou, J., Pitha, P.M., 1996. Sp1 binding plays a critical role in Erb-B2- and v-ras-mediated downregulation of alpha2-integrin expression in human mammary epithelial cells. Mol. Cell. Biol. 16, 6178–6189.

Zhao, L., Chang, L.S., 1997. The human POLD1 gene. Identification of an upstream activator sequence, activation by Sp1 and Sp3, and cell cycle regulation. J. Biol. Chem. 272, 4869–4882.

E

Prodrug-activating systems in suicide gene therapy

Perspective
SERIES
On cancer biotherapy
David H. Kirn, Editor

Caroline J. Springer and Ion Niculescu-Duvaz
Cancer Research Campaign
Centre for Cancer Therapeutics at the Institute of Cancer Research, Sutton, Surrey, United Kingdom

Address correspondence to: Caroline J. Springer, Cancer Research Campaign Centre for Cancer Therapeutics at the Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, United Kingdom.
Phone: 44-208-722-4214; Fax: 44-208-643-6902; E-mail: caroline@icr.ac.uk.

Cancer chemotherapy encompasses a large number of well-documented and clinically established methods for the treatment of malignant diseases. However, the efficacy of these approaches is often hampered by an insufficient therapeutic index, lack of specificity, and the emergence of drug-resistant cell subpopulations. One approach aimed at enhancing the selectivity of cancer chemotherapy for solid tumors relies on the application of gene therapy technologies.

Gene therapies are techniques for modifying the cellular genome for therapeutic benefit. In cancer gene therapy, both malignant and nonmalignant cells may be suitable targets. The possibility of rendering cancer cells more sensitive to chemotherapeutics or toxins by introducing "suicide genes" was suggested in the late 1980s. This approach has two alternatives: toxin gene therapy, in which the genes for toxic products are transfected directly into tumor cells; and enzyme-activating prodrug therapy, in which the transgenes encode enzymes that activate specific prodrugs to create toxic products. The latter approach, known as gene-directed enzyme prodrug therapy (GDEPT) (1, 2) or virus-directed enzyme prodrug therapy (VDEPT) (3), may be used in isolation or combined with other strategies, such as the biotherapies described elsewhere in this Perspective series. VDEPT using selectively replicating viruses as vectors represents a promising means to target suicide genes specifically to tumor cells, an approach that is only beginning to be explored (for examples, see Hermiston, this Perspective series, ref. 4).

GDEPT and VDEPT are two-step treatments for solid tumors (Figure 1). In the first step, the gene for a foreign enzyme is delivered and targeted in a variety of ways to the tumor where it is to be expressed. In the second step, a prodrug is administered that is selectively activated to the drug by the foreign enzyme expressed in the tumor. Ideally, the gene for the enzyme should be expressed exclusively in the tumor cells and should reach a concentration sufficient to activate the prodrug for clinical benefit. The catalytic activity of the expressed protein must be adequate to activate the prodrug under physiological conditions. Because expression of the foreign enzymes will not occur in all cells of a targeted tumor in vivo, a bystander effect (BE) is required, whereby the prodrug is cleaved to an active drug that kills not only the tumor cells in which it is formed, but also neighboring tumor cells that do not express the foreign enzyme (5).

The genes can be engineered to express their products either intracellularly or extracellularly in the recipient cells (6). There are potential advantages to each approach. When the enzyme is intracellularly expressed, the prodrug must enter the cells for activation, and subsequently the active drug must diffuse through the interstitium across the cell membrane to elicit a BE. Cells in which the enzyme is expressed (tethered to the outer surface) are able to activate the prodrug extracellularly. A more substantial BE could therefore be generated with extra-

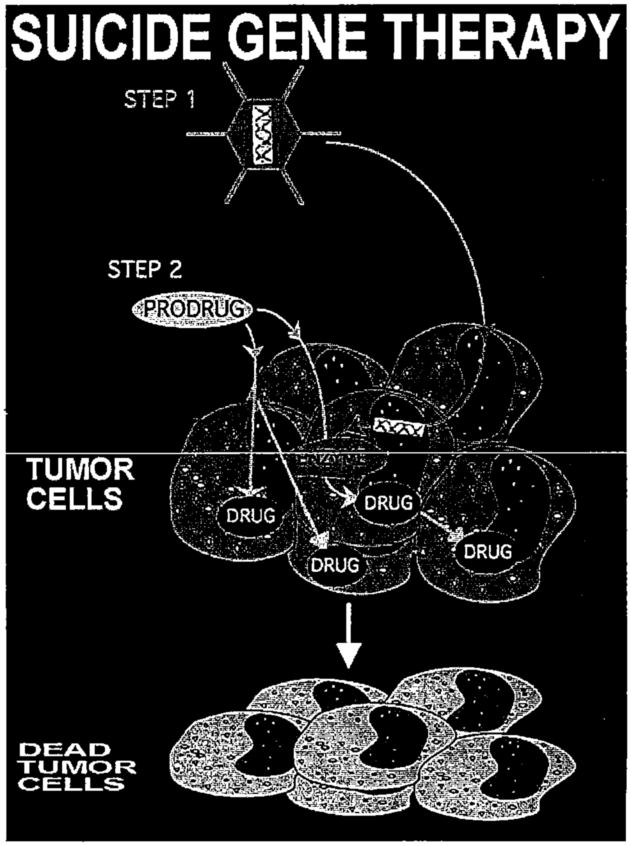


Figure 1
GDEPT, a form of suicide gene therapy, aims to maximize the effect of a toxic drug and minimize its systemic effects by generating the drug in situ within the tumor. In the first step in this procedure, the gene for an exogenous enzyme is delivered and expressed in the tumor cells. Subsequently, a prodrug is administered that is converted to the active drug by the foreign enzyme expressed inside or on the surface of tumor cells.

cellular gene product expression, but spread of the active drug into the general circulation is a possible disadvantage (1, 6).

Parameters that influence the success of GDEPT systems

Effective tumor destruction with GDEPT depends on the design of the gene-therapy vectors, the chemistry of the prodrugs and their toxic metabolites, and the means to deliver one or both components specifically to target cells. Vectors, the vehicles in which the transgenes reach the tumor cell, must be carefully tailored to specific GDEPT systems. The specificity of targeting to cancer cells and efficient transfection are essential for effective GDEPT, as are the toxicity of the vector and the uptake of prodrugs or drugs by normal and malignant cells. GDEPT systems (7–10) require the design of tailored prodrugs and the use of a foreign enzyme — one whose activity is absent from the patient's tissues (or at least from the tissue to be treated), and that can convert the prodrug to the drug in a specific manner. Beneficial immune effects may be induced either by stimulation of the host immune system or by the use of additional cytokine gene therapy (see article on immunomodulation by Agha-Mohammadi and Lotze, this Perspective series, ref. 11). The efficiency of the BE is another key determinant of the success of these systems.

The enzymes proposed for GDEPT fall into two categories. The first comprises foreign enzymes of nonmammalian origin, with or without human homologues. Examples are viral tyrosine kinase (TK), bacterial cytosine deaminase (CD), carboxypeptidase G2 (CPG2), purine nucleoside phosphorylase (PNP), and nitroreductase (NR). The second category consists of enzymes

Table 1
Quantitative data on GDEPT systems

No.	Enzyme/ prodrug system	Potency, IC_{50} (μM)		K_M ($S_{0.5}$) (μM)	V_{max} (nM/mg/min)	Potential of activation[B]	Degree of activation[C]	Clinical trial
		Prodrug	Drug			(fold)	(fold)	
1	CA/CPT-11	1.6–8.1	SN-38: 0.003–0.011	23–52.9	1.43[D]	150–3,000	7–17	1
2	CD/5-FC	26,000	5-FU: 4–23.5	17,900[E] 800[F]	11.7[E,G] 68[F,G]	100–8,000	70–1,000	2
3	CPG2/CMDA, CJS278[H]	CMDA: 1,700–3,125; CJS278: 0.256	CMBA: 8–65; Doxorubicin: 0.012	CMDA: 3.4	CMDA: 583[I] Doxorubicin: 21	CMDA: 26–390; Doxorubicin: 11	CMDA: 10–115;	None
4	Cyt-450/CP, IF, ipomeanol, 2-AA	CP, IF ~4,000	[J]	CP: 300; IF: 480	CP: 39.1; IF: 17.8	[K]	5–60 50–100[L]	1
5	dCK/ara-C	0.3–0.6	[K]	25.6	[K]	[K]	2–100	None
6	HSV-TK/GCV, ACV	GCV: 200–600	GCVTP[K]	GCV: 11–15.8 ACV: 305–375	GCV: 1.3–2.2[M] ACV: 0.3–0.4[M]	[K]	20–1,000	> 21
7	NR/CB1954	> 1,000	0.02[N]	900	6.0[I]	> 50,000	14–10,000	None
8	PNP/6-MePdR	> 200	3.7	14–23[O]	422–638[O,P]	25–1,000	40	None
9	TP/5′-DFUR	17	5-FUdR: 0.0023	325–433	0.17–2.28	7000	165	[K]
10	VZV-TK/ara-M	> 2,000	Ara-MTP < 1[Q]	56	680[R]	> 2,000	55–600	None
11	XGPRT/6-TX, 6-TG	6-TX > 50; 6-TG = 0.5	[K]	[K]	[K]	[K]	6-TX: >20; 6-TG: 10	None

Abbreviations: 2-AA, 2-aminoanthracene; ACV, acyclovir; ara-C, cytosine arabinoside; ara-M, 6-methoxypurine arabinonucleoside; CA, carboxylesterase; CB1954, 5-aziridinyl-2,4-dinitrobenzamide; CD, cytosine deaminase; CMBA, (2-chloroethyl)(2-mesyloxyethyl)aminobenzoic acid; CMDA, (2-chloroethyl)(2-mesyloxyethyl)aminobenzoyl-L-glutamic acid; CP, cyclophosphamide; CPG2, carboxypeptidase G2; CPT-11, irinotecan; cyt-450, cytochrome P450; dCK, deoxycytidine kinase; 5-FC, 5-fluorocytosine; 5′-DFUR, 5′-deoxy-5-fluorouridine; 5-FUdR, 5-fluorodeoxyuridine; 5-FU, 5-fluorouracil; GCV, ganciclovir; GCVTP, ganciclovir triphosphate; HSV-TK, herpes simplex virus thymidine kinase; IF, ifosfamide; 6-MeP, 6-methylpurine; 6-MePdR, 6-methylpurine-2′-deoxyribonucleoside; NR, nitroreductase; PNP, purine nucleoside phosphorylase; SN-38, 7-ethyl-10-hydroxy-camptothecin; 6-TG, 6-thioguanine; TP, thymidine phosphorylase; 6-TX, 6-thioxanthine; VZV-TK, varicella zoster virus thymidine kinase; XGPRT, xanthine-guanine phosphoribosyl transferase. [A]In alphabetical order. [B]Ratio of IC_{50} prodrug/IC_{50} drug in wild-type cell line; [C]ratio of IC_{50} prodrug in wild-type cell line/IC_{50} prodrug in transfected (infected) cell line; [D]pmol/mg/min; [E]bacterial origin; [F]yeast origin; [G]μM/min/μg; [H]N-[4-(L-glutamylcarbonylamino)benzyloxycarbonyl]doxorubicin; [I]s^{-1}; [J]not determined in the same system; [K]no data were obtainable; [L]if cytochrome P450 is coexpressed with P450 reductase; [M]pmol/mg/min; [N]for 5-(aziridine-1-yl)-2-nitro-4-hydroxylamino-benzamide in V79 cells; [O]for inosine, adenine, and guanine nucleosides; [P]μM/min/mg; [Q]inferred from in vitro experiment; [R]relative maximal velocity.

of human origin that are absent or are expressed only at low concentrations in tumor cells, such as deoxycytidine kinase (dCK) and cytochrome P450. The human homologues of enzymes in the first category have different substrate structural requirements than the foreign enzymes have. Their main disadvantage as therapeutic agents may be the potential to elicit an immune response in humans, although this may actually provide benefits to therapy. Enzymes of the second category are unlikely to induce immune responses, but their use will in most cases lead to some prodrug activation in normal cells.

The second step is administration of the prodrug. GDEPT's advantages over conventional prodrug therapy (its greater selectivity and higher drug concentrations) arise from the fact that the drug is generated in situ in the target tumor. Prodrugs for GDEPT must be considerably less cytotoxic than their corresponding active drugs, and must be suitable substrates for the activating enzyme under physiological conditions. In addition, they should be chemically stable under physiological conditions and diffuse readily in the tumor interstitium. They should also have good pharmacological and pharmacokinetic properties, and release an active drug with a good BE. Prodrugs must also be tailored to their site of activation: if activation occurs intracellularly, they must be able to cross the tumor cell membrane; whereas if the enzyme is expressed on the surface of cells, there is no such requirement, but the active drug must be able to cross the cell membrane.

The cytotoxicity differential between the prodrug and its corresponding drug should be as high as possible, and the active drug should be highly diffusible or be actively taken up or exported by cells. The design of a prodrug that can release a highly effective drug requires knowledge of the quantitative structure-activity relationship (QSAR). For this reason, and in order to obtain proof of principle for GDEPT, most of the prodrugs used in suicide gene therapy to date have been clinically licensed anticancer agents with known pharmacological, pharmacokinetic, dosage, and safety parameters. However, it is likely that the next generation of prodrugs will be specially tailored for GDEPT. One possibility is the use of "self-immolative" prodrugs.

A self-immolative prodrug can be defined as a compound that generates an unstable intermediate that then extrudes the active drug in subsequent steps. Although the activation process that generates the unstable species is generally enzyme-mediated, extrusion occurs spontaneously through the fragmentation of the prodrug, often at a distinct site. Self-immolative prodrugs allow their lipophilicity to be altered with minimal effect on the activation kinetics. Indeed, kinetics of activation that are unfavorable due to the chemical or steric features of the active drug can be improved by this approach. The range of drugs that can be converted to prodrugs is greatly extended, and is unrestricted by the structural substrate requirements for a given enzyme. Figure 2 shows one self-immolative prodrug activation reaction, the activation of the doxorubicin prodrug by the enzyme CPG2 (12, 13), followed by the spontaneous extrusion of the DNA-damaging agent doxorubicin.

Table 1 summarizes pharmacokinetic data on 11 enzyme/prodrug pairs that have been designed for use in GDEPT systems. With the exception of cyclophosphamide, ifosfamide, and some prodrugs designed for CPG2 and NR, none of the prodrugs shown are self immolative. Self-immolative prodrugs derived from alkylating agents and anthracyclines have been synthesized for activation by CPG2 (12, 13), and self-immolative derivatives from secocyclopropylindolines and ene-diyne prodrugs have been synthesized for use with NR (7).

Also shown in Table 1 are two parameters that are useful in comparing the different GDEPT systems: the potential of activation of a given system and its degree of activation. The first parameter is defined as the ratio of IC_{50}'s of the prodrug and of the active drug in a nontransfected control tumor cell line. It represents the maximum possible efficiency of a given enzyme/prodrug system toward that cell line. The degree of activa-

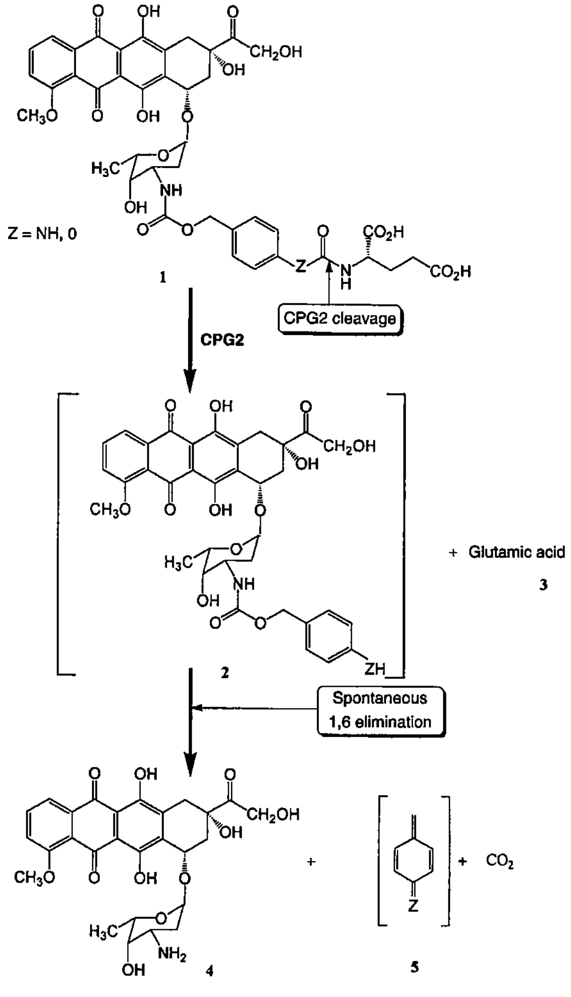


Figure 2

Self immolation of prodrug1 to yield the chemotherapeutic drug doxorubicin. The doxorubicin prodrug (**1**) is cleaved by carboxypeptidase G2 (CPG2), releasing the glutamic acid (**3**) and an unstable intermediate (**2**). The latter undergoes a spontaneous 1,6 elimination, extruding doxorubicin (**4**), a quinone imine (**5**), and carbon dioxide. This scheme can be readily modified to allow the production of structurally similar drugs, such as daunorubicin.

tion is defined as the ratio of the IC_{50} of the prodrug in the nontransfected cell line to the IC_{50} of the prodrug in the transfected or infected derivative of the cell line that expresses the activating enzyme. Both the potential for activation and the degree of activation depend on the cell line's sensitivity to the drug, but the degree of activation also reflects the efficiency of the prodrug system in the context of that cell line; if the enzyme in the transfected line is sufficient to convert the prodrug immediately and quantitatively to the active form of the drug, the degree of activation is identical to the potential of activation for that GDEPT system.

Improving activation kinetics

The concentration of the drug and the rate at which it is released at the activation site depend on the kinetic parameters of the enzyme/prodrug system. Unfortunately, the published V_{max} and K_m values – which measure the maximum velocity of the activation reaction and the concentration of substrate needed to reach half this maximum velocity – are not sufficient to allow enzyme/prodrug systems to be compared. As a rule, however, low K_m and high V_{max} (or k_{cat}) would be expected to be found in relatively effective systems, and the comparison of the yeast CD with bacterial CD bears out this prediction. As shown in Table 1, the yeast enzyme, which proved to be more effective than its bacterial counterpart in a GDEPT experiment, exhibits lower K_m and higher V_{max} values (14). However, the literature does not supply consistent values for the V_{max} of these enzymes, because V_{max} has been determined under very different experimental conditions for the various systems and is expressed in different ways, making direct comparisons impossible. Despite these caveats, it appears from the data in Table 1 that prodrugs such as CMDA (a substrate of CPG2), GCV (a substrate of HSV-TK), and CPT-11 (a substrate of CA) are superior to 5-FC (a substrate of CD) or 5′-FDUR (a substrate of TP), because the latter two have high K_m and low V_{max} values. The turnover number, k_{cat}, provides additional information about the reaction rate, but the implications of this measure for tumor cell killing is unclear, because it is not yet known if sudden release of the active drug is more effective than slow, constant release, or if quiescent and proliferating cells differ in their sensitivity to drugs released at different rates.

New techniques that are available to increase the efficacy of enzymes to activate prodrugs for GDEPT were recently reviewed (9). Some of these approaches build on crystallographic descriptions of the active site of the enzyme involved in the enzyme/prodrug system, which should permit the molecular modeling, and eventually the rational synthesis, of substrates that are well-suited for a GDEPT system. An alternative is to modify the active site of the enzyme by site-directed mutagenesis in order to increase its catalytic efficiency toward an existing substrate. Black et al. (15) applied these techniques to obtain mutants of HSV-TK with improved kinetic parameters for the prodrugs GCV and ACV. Similarly, Smith and colleagues (16) performed site-directed mutagenesis on carboxypeptidase A to improve the efficiency of this enzyme toward specific substrates that, by design, are less prone to interfere with other human peptidases (16).

Prodrugs may also be activated by a metabolic cascade involving the sequential action of several enzymes, for example the activation of GCV to GCV triphosphate (GCVTP) by three different kinases (HSV-TK, guanylate kinase, and nucleoside diphosphate kinase), acting in series. This approach requires the cotransfection of genes for each of the enzymes, but is expected to increase the overall yield of the desired final metabolite, the active drug. In the case of GCV, Blanche et al. recently claimed that the simultaneous transfection of these three genes allows cells to convert more than 90% of the prodrug to GCVTP (17). Likewise, the cotransfection of the genes for cytochrome P450 and P450 reductase significantly increases the conversion of CP to it toxic metabolites, and therefore improves the overall efficiency of the cytochrome P450/CP GDEPT system (18) (see Table 1).

Interpreting potential of activation and degree of activation

As discussed above, the potential of activation of a GDEPT system reflects its maximal theoretical efficiency, at least toward a specific cell type. Unfortunately, not all the systems can be defined in this way, because multiple products may be released, and the toxicity of each of these metabolites may not be available. Thus, although GCV is relatively nontoxic, its monophosphorylated derivative, GVCMP, is highly cytotoxic (19), so the potential of activation of this system cannot be calculated accurately from the known IC_{50} of the triphosphorylated derivative, GCVTP. Likewise, for CP activated by cytochrome P450, a maximum differential of 20- to 25-fold is theoretically achievable based on the IC_{50} values of CP and its corresponding phosphoramide mustard. However, the degree of activation obtained was found to be 100-fold (20), which suggests that the CP phosphoramide mustard is not the final active metabolite.

The optimal approach to improving the activation potential of a system is to design prodrugs with lower cytotoxicity, but the complementary strategies of increasing the cytotoxicity or improving the efficiency of release of the drug are also helpful. Some highly cytotoxic compounds (with IC_{50} in the nM range) such as ene-diynes, cyclopropylindolines, and taxoids are now available, but generally their structures are complicated and efficient ways are needed to convert them to low-cytotoxicity prodrugs. Designing self-immolative analogues of these prodrugs could be a way to move forward, as could modifications that improve the uptake of the prodrug by enzymatic modifications (21) or alter the lipophilicity of the prodrug (22). The latter capability is especially useful for tailoring the prodrug for use with an extracellular or intracellular activating enzyme.

The degree of activation, which reflects the efficiency of the system, is another parameter useful for characterizing a GDEPT system. By definition, it must be lower than or at least equal to the potential of activation for the system, as is seen for all the systems analyzed in Table 1. The interpretation of both of these parameters is complex. Their values offer some insight into the in vitro situation, where a single cell type, transfected with a gene for an activating enzyme, is challenged with the prodrug or its toxic metabolites either before or after transfection. For several reasons, these values may not accurately reflect the situation in vivo. Additional factors, such as pharmacokinetics, prodrug distribution, and immune responses complicate the overall picture. Moreover, obtained IC_{50} values may vary for different cell types, and not all cells in a tumor may be accurately modeled by the cell line chosen for in vitro study. Despite these concerns, these parameters provide a rational basis for comparing different GDEPT systems and should also be helpful in designing new systems.

Increasing the BE

The extent of the BE can be determined from the effect of the treatment on non-genetically modified cells that takes place after prodrug administration, when only a fraction of the tumor mass is genetically modified to express an activating enzyme (5, 8). The striking successes described in GDEPT would surely not be possible in the absence of such an effect. As described, some models require only 1-2% of cells to be genetically modified to obtain therapeutically significant results.

Toxic metabolites are formed after prodrug activation and may be released from dead and dying genetically modified cells. This mechanism is postulated for 5-FU formed from 5-FC; for the metabolites of CP and IP, aldophosphamide, phosphoramidic mustards, and acrolein; for benzoic acid mustard released from CMDA; and for 6-MeP, formed from the corresponding deoxynucleoside. Supporting this model for the BE is the observation that cell-to-cell contact is not required for the killing of untransfected cells by these agents, either in vitro or in vivo. In vitro, 30% of cells expressing CD suffices for the eradication of a whole cell population by 5-FC. This BE is dramatically greater in vivo, even with immunocompromised mice: 2% CD^+ tumor cells can yield 100% tumor regression in athymic mice; with 4% CD^+ cells, 66% of animals are cured of their xenografts. When similar experiments are performed in immunocompetent animals, the results are better still.

For purine or pyrimidine nucleosides, the toxic metabolites are not diffusible across cell membranes, so the HSV-TK/GCV system apparently requires cell-to-cell contact, specifically gap junction formation (23), to display a BE. Consistent with this model, one report showed that tumor cells resistant to BE did not show dye transfer from cell to cell, whereas BE-sensitive tumor cells did. Furthermore, dieldrin, a drug known to decrease gap junction communications, diminished the dye transfer and also inhibited the BE, leading to the suggestion that the BE of this system could be enhanced by pharmacological manipulation of the gap junctions in vivo. Apigenin, a flavonoid, and lovastatin, an inhibitor of HMG-CoA reductase, both upregulate gap junction function and dye transfer in tumors expressing gap junctions. Touraine et al. (24) studied the control of tumors grown from a mixture of 10% $HSV\text{-}TK^+$ adenocarcinoma cells and 90% TK^- cells using GCV. In the absence of lovastatin or apigenin, 30% of animals treated become tumor-free, but when tumor-bearing mice were administered lovastatin or apigenin during GCV treatment, their antitumor response rate doubled.

Although these results are consistent with the hypothesis that gap junctions mediate the BE after GCV treatment, other data suggest that additional mechanisms are involved. In one study of human lung tumor cell lines of different origins, significant cell killing occurred when only 10% of cells expressed HSV-TK. In this system, gap junction communication was not apparent from measuring the rapid intercellular transport of Lucifer Yellow, which detects "rapid-transfer" gap junctional communications, although it could be seen by the slow transfer of a different dye, calcein-AM, which measures the "slow-transfer" gap junctions. However, neither an inhibitor (1-octanol) nor an enhancer (all-*trans* retinoic acid) of gap junction communication affected the extent of the BE, suggesting either that low levels of gap junctions can produce a maximal BE or that bystander cell killing occurs by other means (25). Boucher et al. compared the efficacy of the HSV-TK/GCV system in two human carcinoma cell lines after exposure to GCV and found that the BE depended on the concentration of the enzyme, the number of cells expressing HSV-TK, and the overall confluence of the cells, but not on the activity of functional gap junctions, as assessed by the Lucifer Yellow assay (26).

Another suggestion is that the TK enzyme is transported by apoptotic vesicles or through gap junctions. Phagocytosis of material (e.g., hydrolases or other lytic enzymes) from dying TK^+ cells by bystander cells has also been suggested as a mechanism for the BE. Apoptosis was detected in bystander cells and it was found that this event could be inhibited by BCL2 expression. However, during the apoptosis induction period in bystander cells cocultured with HSV-TK-expressing cells, no phagocytosis was observed. It has also been suggested that killing of tumor cells by apoptosis could heighten the immune response to wild-type tumor cells by a priming effect.

A quantitative expression of the BE was recently proposed using the NR/CB1954 system in a range of human tumor-cell types. The IC_{50}'s of non-NR-expressing cells were measured in the presence of varying proportions of NR-expressing cells. The shift in IC_{50} was used to calculate a value for the BE, termed the transmission efficiency (TE), which is the decrease in IC_{50} caused by the BE, expressed as a percentage of the maximum measured decrease. The percentage of NR-expressing cells for which the TE was 50% (TE_{50}) is a single data point for the BE. The TE_{50} in the cell lines ranged from 0.3% to approximately 2% (27).

There were early suggestions that the immune response improves the efficacy of GDEPT. Although the BE has been observed in immunocompromised animals, recent findings suggest that the BE in vivo is mediated largely through the release of cytokines. Ramesh et al. (28) report that GCV treatment of carcinomas that contained a mixture of $HSV\text{-}TK^+$ cells resulted in almost total tumor ablation in immunocompetent mice, but not in athymic animals of the same strain. In a similar experiment, when *HSV-TK* was transfected into cells grown as xenografts, the tumor growth was inhibited for up to 50 days in GCV-treated, immunocompromised nude mice, but failed to eliminate all the tumor cells in these animals, and tumors regrew 40–50 days after implantation. By contrast, immunocompetent BALB/c mice developed long-lasting immunity in response to *HSV-TK* transduction followed by GCV treatment (29). Taken together, these studies strongly suggest that an intact immune system is important for long-term tumor suppression with TK in vivo.

IL-2 appears to be critical for immune-mediated tumor suppression in this system. In one experiment, cells grown as xenografts in syngeneic mice were injected with an adenoviral vector containing the HSV *TK* gene, the *IL2* gene, or both, followed by treatment with GCV (30). Whereas the tumors continued to grow in the animals injected with a control vector or the vector carrying *IL2*, those treated with HSV TK, with or without coadministration of *IL2*, exhibited tumor necrosis and regression. However, only animals treated with both genes developed effective systemic antitumoral immunity against tumorigenic rechallenges. The antitumoral immunity was associated with the presence of tumor-specific cytolytic $CD8^+$ T lymphocytes. A third vector containing the mouse *GMCSF* gene enhanced and prolonged this antitumoral immunity, allowing animals treated with all three genes to survive for longer than four months without recurrence. These and similar findings (30) establish the synergism between suicide gene and cytokine gene therapies.

Future perspectives

Some hurdles must be overcome before GDEPT will become a clinically efficient treatment of cancers. Major improvements are needed in vector design to enhance targeting and delivery of suicide genes. Multiple options are available, including nonviral vectors, more complex systems involving coexpression of suicide genes with immunological or tumor-suppressor genes, and selectively replicating viruses. Double suicide gene therapy, in which a combination of suicide genes is introduced simultaneously, shows promise in vitro and in vivo. The released active drugs in such an approach can act by different mechanisms, leading to a synergistic effect on tumor cells or an enhanced BE, particularly if cell-permeant and cell-impermeant active metabolites can be released together. Additionally, the occurrence of resistant populations is less likely for drugs with different mechanisms of action. Uckert et al. (31) have shown for human carcinoma cell lines grown in vivo that double suicide gene therapy (involving *HSV-TK* and *CD*) allowed the elimination of tumors, but neither gene applied individually gave this result.

Thus, GDEPT systems have already shown efficacy in vivo. Future developments in this technology should use mutagenesis to obtain more efficient activation of a given prodrug, or to adapt the active site so that it binds better to prodrugs that are not substrates for endogenous enzymes. The prodrugs, too, should be redesigned to create better substrates for the enzymes, to maximize drug release or the BE, to take advantage of self-immolative strategies of activation, or to allow the active drug to accumulate more readily in tumor cells. Finally, it will also be useful to investigate the ways in which different prodrug systems synergize with each other or with other cancer treatments. The combination of GDEPT with radiotherapy or immunotherapy has previously been suggested. Such approaches may involve a sequential treatment schedule (GDEPT/radiation therapy or GDEPT/immunotherapy). The transfection of suicide genes together with genes that are able to increase the sensitivity of the tumors to radiation or enhance the potential of the host immune system is an alternative strategy. Other combination therapies are possible, such as applying GDEPT in conjunction with replicating oncolytic adenoviruses, such as ONYX-015 (32; see also Heise and Kirn, this Perspective series, ref. 33) or ONYX-838, or the use of such viruses as carriers for the suicide genes (see Hermiston, this Perspective series, ref. 4). The combination of these replicating adenoviruses with conventional chemotherapy has proven highly effective, and replacing the chemotherapeutic arm in the ONYX-015 study with GDEPT might well provide additional benefits.

Acknowledgments

This work was funded by the Cancer Research Campaign (grants SP2330/0201 and SP2330/0102) and the Institute of Cancer Research. We would like to thank P. Workman and K. Harrap for support.

1. Marais, R., Spooner, R.A., Light, Y., Martin, J., and Springer, C.J. 1996. Gene-directed enzyme prodrug therapy with a mustard prodrug/carboxypeptidase G2 combination. *Cancer Res.* **56**:4735–4742.
2. Bridgewater, J.A., et al. 1995. Expression of the bacterial nitroreductase enzyme in mammalian cells renders them selectively sensitive to killing by the prodrug CB1954. *Eur. J. Cancer.* **31A**:2362–2370.
3. Huber, B.E., Richards, C.A., and Austin, E.A. 1995. VDEPT: an enzyme/prodrug gene therapy approach for the treatment of metastatic colorectal cancer. *Advanced Drug Delivery Reviews.* **17**:279–292.
4. Hermiston, T. 2000. Gene delivery from replication-selective viruses: arming guided missiles in the war against cancer. *J. Clin. Invest.* **105**:1169–1172.
5. Huber, B.E., Austin, E.A., Richards, C.A., Davis, S.T., and Good, S.S. 1994. Metabolism of 5-fluorocytidine to 5-fluorouracil in human colorectal tumor cells transduced with the cytosine deaminase gene: significant antitumor effects when only a small percentage of tumor cells express cytosine deaminase. *Proc. Natl. Acad. Sci. USA.* **91**:8302–8306.
6. Marais, R., et al. 1997. A cell surface tethered enzyme improves efficiency in gene-directed enzyme prodrug therapy. *Nat. Biotechnol.* **15**:1373–1377.
7. Denny, W.A., and Wilson, W.R. 1998. The design of selectively-activated anti-cancer prodrugs for use in antibody-directed and gene-directed enzyme prodrug therapies. *J. Pharm. Pharmacol.* **50**:387–394.
8. Niculescu-Duvaz, I., Spooner, R., Marais, R., and Springer, C.J. 1998. Gene-directed enzyme prodrug therapy. *Bioconjug. Chem.* **9**:4–22.
9. Encell, L.P., Landis, D.M., and Loeb, L.A. 1999. Improving enzymes for gene therapy. *Nat. Biotechnol.* **17**:143–147.
10. Springer, C.J., and Niculescu-Duvaz, I. 1999. Patent property of prodrug involving gene therapy (1996-1999). *Expert Opinion on Therapeutical Patents.* **9**:1381–1388.
11. Agha-Mohammadi, S., and Lotze, M.T. 2000. Immunomodulation of cancer: potential use of selectively replicating agents. *J. Clin. Invest.* **105**:1173–1176.
12. Niculescu-Duvaz, D., et al. 1998. Self-immolative nitrogen mustard prodrugs for suicide gene therapy. *J. Med. Chem.* **41**:5297–5309.
13. Niculescu-Duvaz, I., et al. 1999. Self-immolative anthracycline prodrugs for suicide gene therapy. *J. Med. Chem.* **42**:2485–2489.
14. Kievit, E., et al. 1999. Superiority of yeast over bacterial cytosine deaminase for enzyme/prodrug gene therapy in colon cancer. *Cancer Res.* **59**:1417–1421.

15. Black, M.E., Newcomb, T.G., Wilson, H.M., and Loeb, L.A. 1996. Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proc. Natl. Acad. Sci. USA.* **93**:3525–3529.
16. Smith, G.K., et al. 1997. Toward antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1 and novel in vivo stable prodrugs of methotrexate. *J. Biol. Chem.* **272**:15804–15816.
17. Blanche, F., Cameron, B., Couder, M., and Crouzet, J. 1997. Rhone-Poulenc Rorer. Enzymes combinations for destroying proliferative cells. Patent WO 9735024
18. Chen, L., Yu, L.J., and Waxman, D.J. 1997. Potentiation of cytochrome P450/cyclophosphamide-based cancer gene therapy by coexpression of the *P450 reductase* gene. *Cancer Res.* **57**:4830–4837.
19. Rubsam, L.Z., Davidson, L., and Shewach, D.S. 1998. Superior cytotoxicity with ganciclovir compared with acyclovir and 1-beta-D-arabinofuranosylthymine in herpes simplex virus-thymidine kinase-expressing-cells: a novel paradigm for cell killing. *Cancer Res.* **58**:3873–3882.
20. Waxman, D.J., and Goldstein, J.A. 1999. Trustee of Boston University. Methods of using cytochrome P450 reductase for the enhancement of P450-based anti-cancer gene therapy. Patent WO 9905299.
21. Tiraby, M., et al. 1998. Concomitant expression of *E. coli* cytosine deaminase and uracil phosphoribosyltransferase improves the cytotoxicity of 5-fluorocytosine. *FEMS Microbiol. Lett.* **167**:41–49.
22. Balzarini, G., et al. 1998. Superior cytostatic activity of the ganciclovir elaidic acid ester due to the prolonged intracellular retention of ganciclovir anabolites in herpes simplex virus type 1 thymidine kinase gene-transfected tumor cells. *Gene Ther.* **5**:419–426.
23. Touraine, R.L., Ishii-Morita, H., Ramsey, W.J., and Blaese, R.M. 1998. The bystander effect in the HSVtk/ganciclovir system and its relation to gap junctional communication. *Gene Ther.* **5**:1705–1711.
24. Touraine, R.L., Vahanian, N., Ramsey, W.J., and Blaese, R.M. 1998. Enhancement of the herpes simplex virus thymidine kinase/ganciclovir bystander effect and its antitumor efficacy in vivo by pharmacologic manipulation of gap junctions. *Hum. Gene Ther.* **9**:2385–2391.
25. Imaizumi, K., et al. 1998. Bystander tumoricidal effect and gap junctional communication in lung cancer cells. *Am. J. Respir. Cell Mol. Biol.* **18**:205–212.
26. Boucher, P.D., Ruch, R.J., and Shewach, D.S. 1998. Differential ganciclovir-mediated cytotoxicity and bystander killing in human colon carcinoma cell lines expressing herpes simplex virus thymidine kinase. *Hum. Gene Ther.* **9**:801–814.
27. Friedlos, F., Court, S., Ford, M., Denny, W.A., and Springer, C.J. 1998. Gene-directed enzyme prodrug therapy: quantitative bystander cytotoxicity and DNA damage induced by CB1954 in cells expressing bacterial nitroreductase. *Gene Ther.* **5**:105–112.
28. Ramesh, R., Marrogi, A.J., Munshi, A., Abboud, C.N., and Freeman, S.M. 1996. In vivo analysis of the "bystander effect": a cytokine cascade. *Exp. Hematol.* **24**:829–838.
29. Pavlovic, J., Nawrath, M., Tu, R., Heinicke, T., and Moelling, K. 1996. Antitumor immunity is involved in the thymidine kinase-mediated killing of tumors induced by activated Ki-ras (G12V). *Gene Ther.* **3**:635–643.
30. Chen, S.H., et al. 1996. Combination suicide and cytokine gene therapy for hepatic metastases of colon carcinoma: sustained antitumor immunity prolongs animal survival. *Cancer Res.* **56**:3758–3762.
31. Uckert, W., et al. 1998. Double suicide gene (cytosine deaminase and herpes simplex virus thymidine kinase) but not single gene transfer allows reliable elimination of tumor cells in vivo. *Hum. Gene Ther.* **9**:855–865.
33. Heise, C.C., Williams, A., Olesch, J., and Kirn, D.H. 1999. Efficacy of a replication-competent adenovirus (ONYX-015) after intratumoral injection: intratumoral spread and distribution effects. *Cancer Gene Ther.* **6**:499–504.
32. Heise, C., and Kirn, D.H. 2000. Replication-selective adenoviruses as oncolytic agents. *J. Clin. Invest.* **105**:847–851.